

FIUBA - 75.06

Organización de Datos

Trabajo práctico 1, Grupo 28
1er cuatrimestre, 2019

Integrantes:

Nombre	Padrón
GAIDO, Nicolás	100856
MARCÓ DEL PONT, Matías	101302
ROMERO VÁZQUEZ, Maximiliano	99118

Fecha de Entrega: 22/04/2019

Link de GitHub:

<https://github.com/MatiasMdelP/Organizacion-de-Datos-TP1>

Índice

1. Introducción	2
2. Objetivo	2
3. Hipótesis	2
4. Análisis	2
4.1. Auctions	2
4.1.1. Cómo varía la cantidad de subastas según los días	3
4.1.2. A qué hora se realizan la mayor cantidad de subastas	4
4.1.3. Cómo se distribuye cantidad de subastas según el día y la hora	5
4.1.4. Cómo varía según las características del dispositivo	6
4.1.5. Como se distribuye la cantidad de subastas en las que participa cada dispositivo	6
4.1.6. Cuanto tiempo tarda en aparecer un mismo dispositivo en una subasta	7
4.1.7. Existe alguna tendencia sobre la fuente de donde procede la subasta	8
4.2. Clicks	8
4.2.1. Cómo evoluciona en el tiempo la cantidad de clicks	9
4.2.2. A qué hora se realizan la mayor cantidad de clicks	10
4.2.3. En que parte de la pantalla los usuarios hacen click	11
4.2.4. Cuáles compañías telefónicas son las más usadas	12
4.2.5. Cuanto tardan los usuarios en hacer click	13
4.2.6. Que marcas de celular registran más clicks	15
4.2.7. Que usuarios realizan mayor cantidad de clicks	16
4.2.8. Otros datos	16
4.3. Events	16
4.3.1. Cómo varía la cantidad de eventos acorde a las variables temporales	17
4.3.2. Cómo varía según el tipo y características de evento	18
4.3.3. Cómo varía la cantidad de eventos acorde a las variables geográficas	20
4.3.4. Cómo varía según las características del dispositivo	21
4.3.5. Cómo varía según tipo de conexión	22
4.3.6. Cómo varía el lapso de tiempo entre dos eventos consecutivos para el mismo usuario	25
4.4. Installs	25
4.4.1. Cómo varía la cantidad de instalaciones acorde a los días	26
4.4.2. A qué hora se realizan la mayor cantidad de instalaciones	27
4.4.3. Cómo se distribuye cantidad de instalaciones según el día y la hora	28
4.4.4. Cómo varía según las características del dispositivo	29
4.4.5. Cómo se distribuye la cantidad de instalaciones por usuario	30
4.4.6. Cómo varía según tipo de conexión	31
4.4.7. Como se distribuyen las instalaciones según la aplicación	31
4.4.8. Como varían las instalaciones según el anunciante	32
4.4.9. Existe alguna tendencia en la forma de instalación	32
5. Conclusión	33

1. Introducción

Este trabajo práctico consiste en realizar el análisis exploratorio del set de datos provisto por la empresa **Jampp**. El que cuenta con cuatro archivos distintos en formato 'csv' (*auctions*, *clicks*, *events* e *installs*).

2. Objetivo

El objetivo del informe es poder realizar un aporte a Jampp que contenga información relevante de los datos presentes en los diferentes archivos, como así también, identificar *patrones* o *funnels* de usuarios. Es decir, tratar de observar comportamientos específicos de los usuarios a los cuales se les muestra una publicidad y ellos conviertan o generen un evento dentro de la aplicación.

3. Hipótesis

Algunas hipótesis manejadas antes de la realización del análisis exploratorio de los datos son:

- El menor flujo de datos se da a la hora de descanso de los usuarios, por lo que entre las 3 am y las 8 am tendríamos que tener una cantidad menor de datos.
- El mayor flujo de datos, ya sean subastas, clicks y demás, lo tendremos los días Viernes, Sábado y Domingo.
- Para realizar los diferentes eventos y subastas, el usuario lo hace a través de una conexión WiFi. Ya que actualmente la mayoría de los lugares (ya sean casas, oficinas, bares, etc.) cuentan con dicha conexión y no requieren de algún gasto con la operadora telefónica.
- Dentro del archivo en donde se encuentra la información referida a las instalaciones tendremos una cantidad mucho menor con respecto a los demás archivos. Debido a la que gente no suele instalar mucho las aplicaciones, y si lo hace, la realiza una vez sola. Lo que varía con respecto a la cantidad de subastas en las que puede participar, o de los eventos generados.

4. Análisis

Luego del análisis de datos correspondiente de todos el set de datos, podemos confirmar que se tratan de datos que comprendidos en el período que va de 5 de Marzo al 13 de Marzo de 2019, y corresponden a Uruguay.

4.1. Auctions

A continuación se analizarán los datos presentados en el archivo "*auctions.csv*" que contiene información acerca de las distintas subastas en las que puede participar **Jampp**. Es decir, son aquellos espacios en los cuales se puede ofrecer un dinero a cambio de mostrar una publicidad en el espacio de una aplicación.

4.1.1. Cómo varía la cantidad de subastas según los días

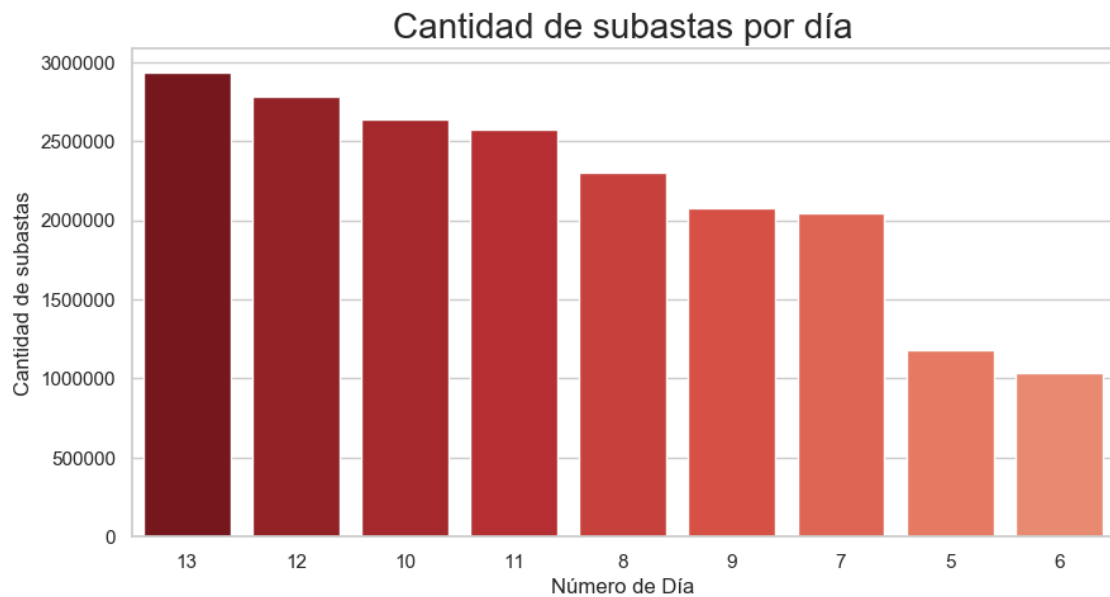


Figura 1: Cantidad de subastas registradas según el día.

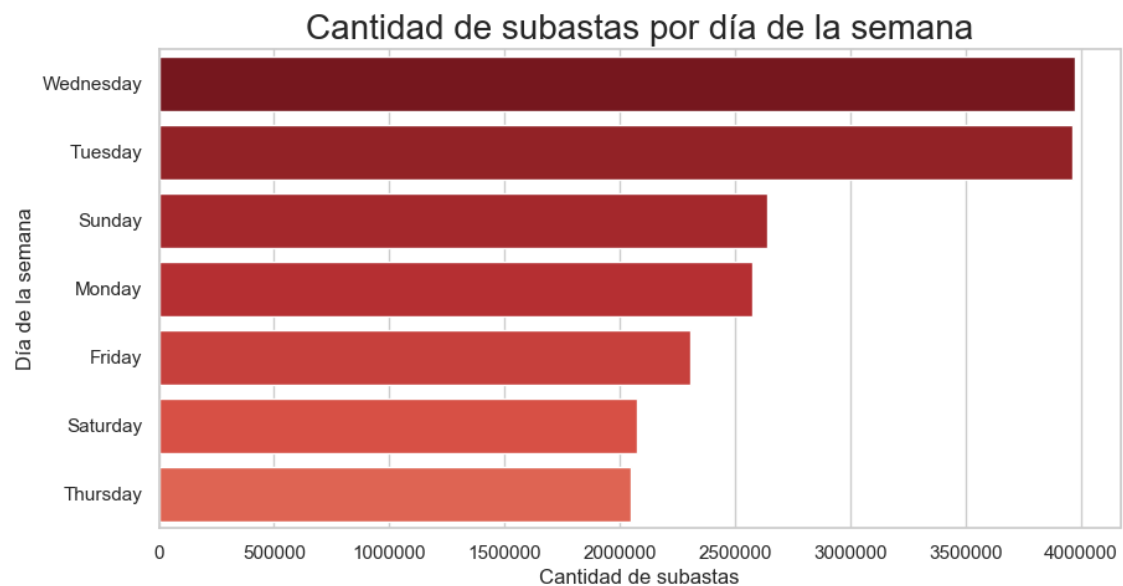


Figura 2: Cantidad de subastas registradas según el día de la semana.

Como podemos ver en los dos gráficos anteriores, la mayor cantidad de subastas se dan en los últimos días analizados. En este caso serían 13, 12, 10 y 11 de Marzo de 2019. Que corresponden a los días Miércoles y Martes.

4.1.2. A qué hora se realizan la mayor cantidad de subastas

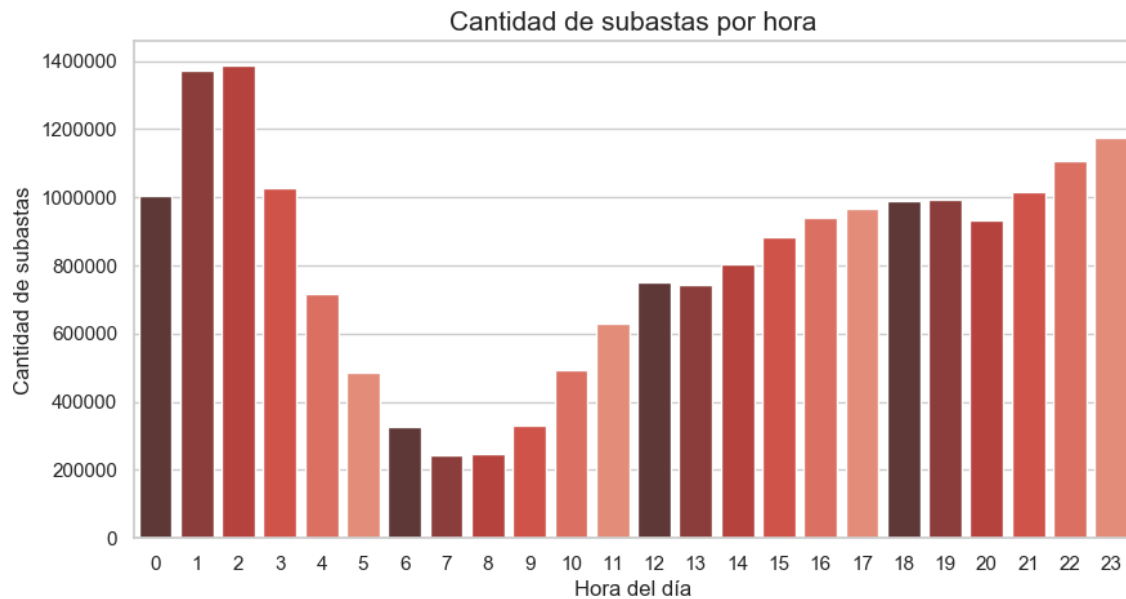


Figura 3: Cantidad de subastas registradas según la hora.

Del anterior gráficos podemos ver:

- La mayor cantidad de subastas se da a las 2 de la madrugada. Y también el mayor flujo se da, entre llegando a la medianoche del día y las primeras horas de la madrugada del día siguiente.
- La menor cantidad se da entre el intervalo de las 6 y las 9 de la mañana. En cada una de éstas horas podemos ver que la cantidad de subastas es inferior a 400.000.

4.1.3. Cómo se distribuye cantidad de subastas según el día y la hora

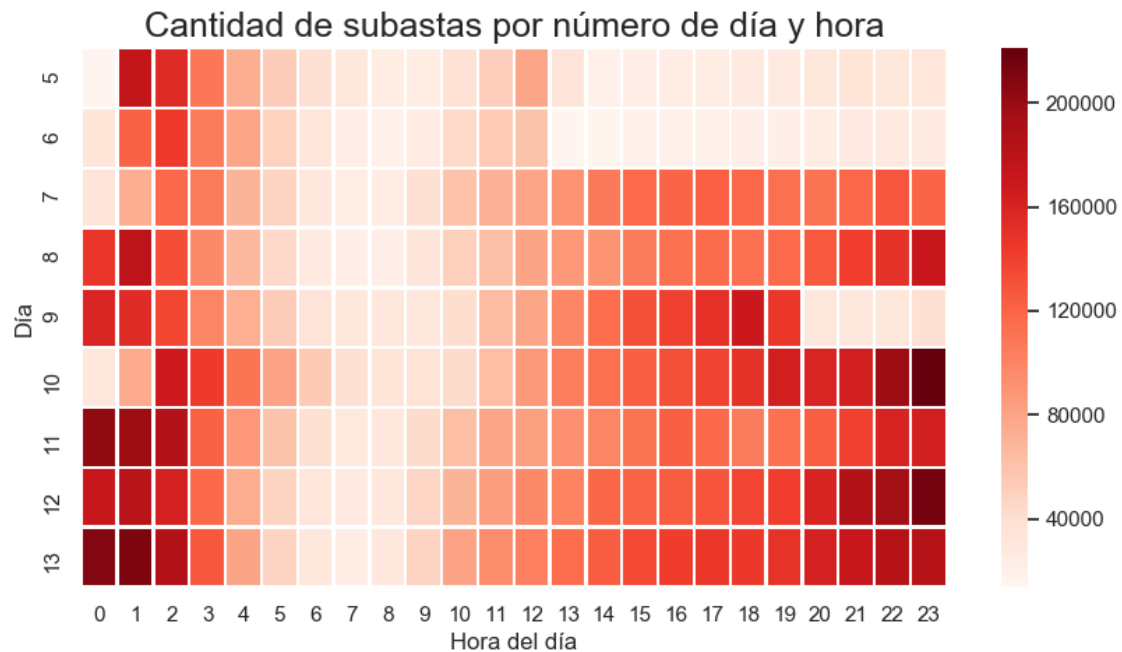


Figura 4: Subastas registradas según el día y la hora.

A partir del gráfico podemos confirmar que, sin importar el día, entre las 6 y las 9 de la mañana es donde se produce la menor cantidad de subastas. Esto suponemos que se trata debido a la hora de descanso de los distintos usuarios.

Por otro lado, nos llama la atención la poca cantidad de subastas que se producen a partir del mediodía de los días 5 y 6 del mes de Marzo. Registrándose un mayor cantidad en la madrugada de dichos días.

4.1.4. Cómo varía según las características del dispositivo

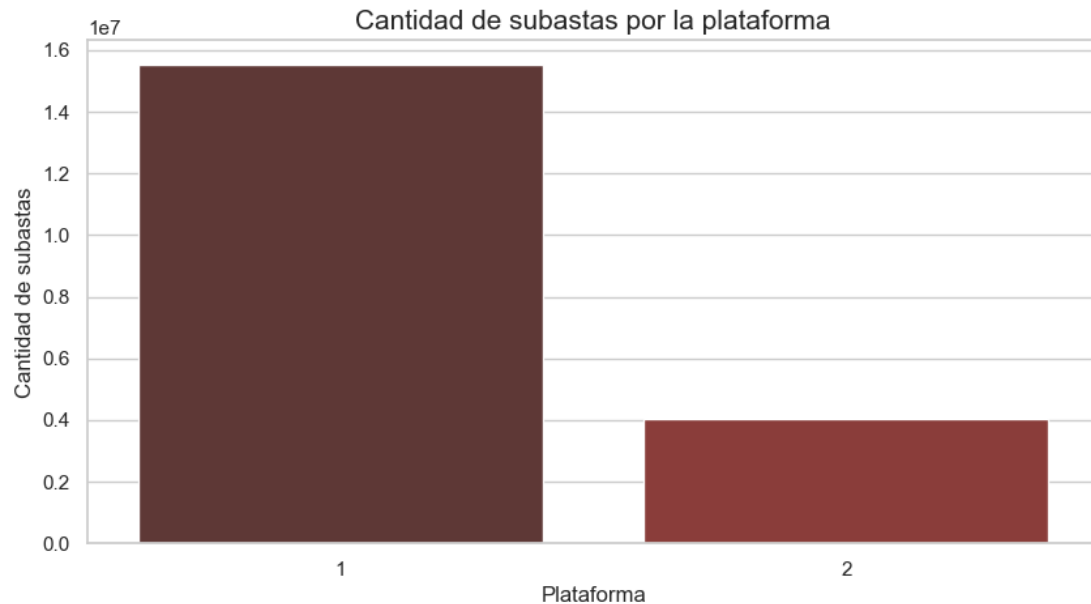


Figura 5: Subastas registradas según la plataforma.

Analizando la plataforma, podemos observar una marcada diferencia hacia aquella con representada con el 1. Si tenemos que arriesgar, diríamos que se trata de la plataforma *Android*.

4.1.5. Como se distribuye la cantidad de subastas en las que participa cada dispositivo

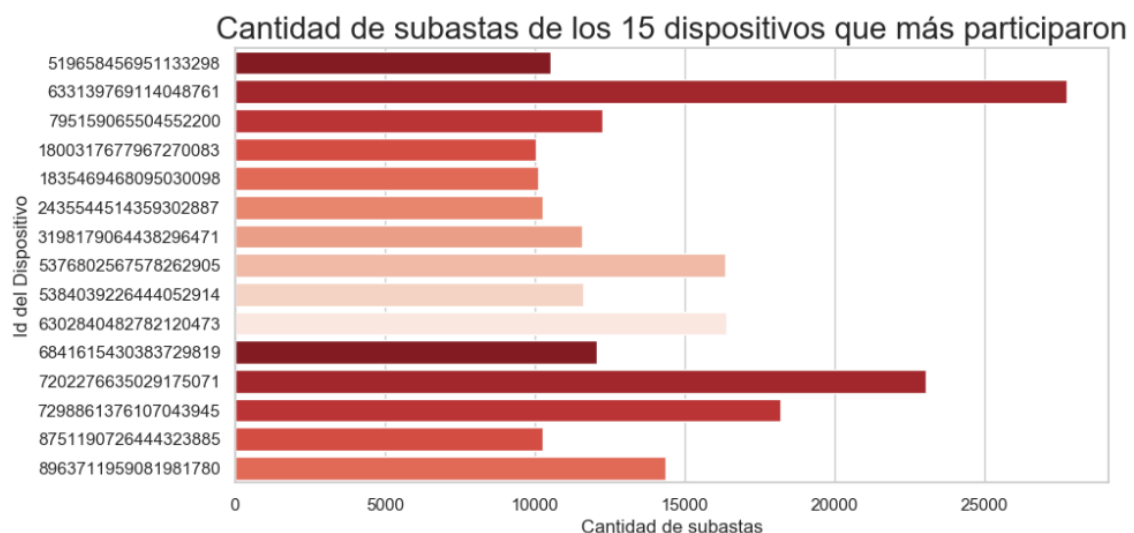


Figura 6: Top 15 de dispositivos con más participación en subastas.

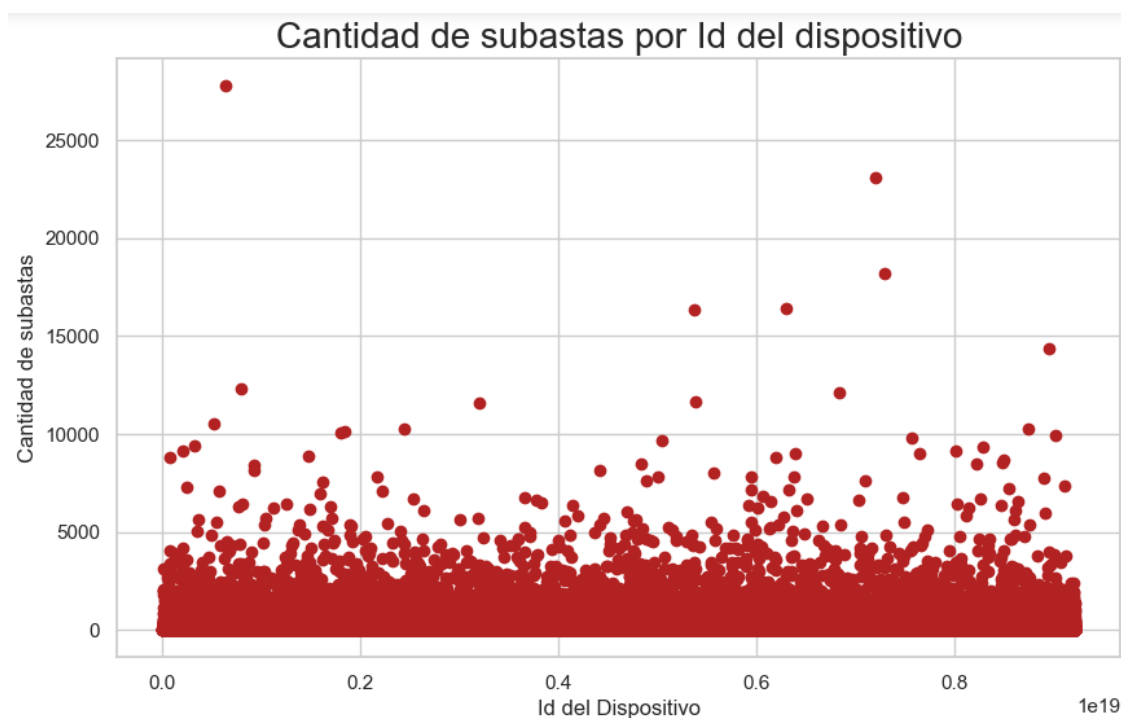


Figura 7: Subastas por ID del dispositivo.

Como se puede ver en los dos gráficos anteriores (Figuras 6 y 7), tenemos que varios dispositivos se encuentran bastante por encima de la media, que indica que cada dispositivo móvil participa entre 94 y 95 subastas en el período del 5 de Marzo al 13 de Marzo de 2019. Siendo estos, valores aislados con respecto a todos los demás usuarios.

El dispositivo que más participa en subastas lo hace con una cantidad de 27.762 veces. Lo que es un valor extraño ya que indicaría que participa en casi 2 subastas por minuto.

4.1.6. Cuanto tiempo tarda en aparecer un mismo dispositivo en una subasta

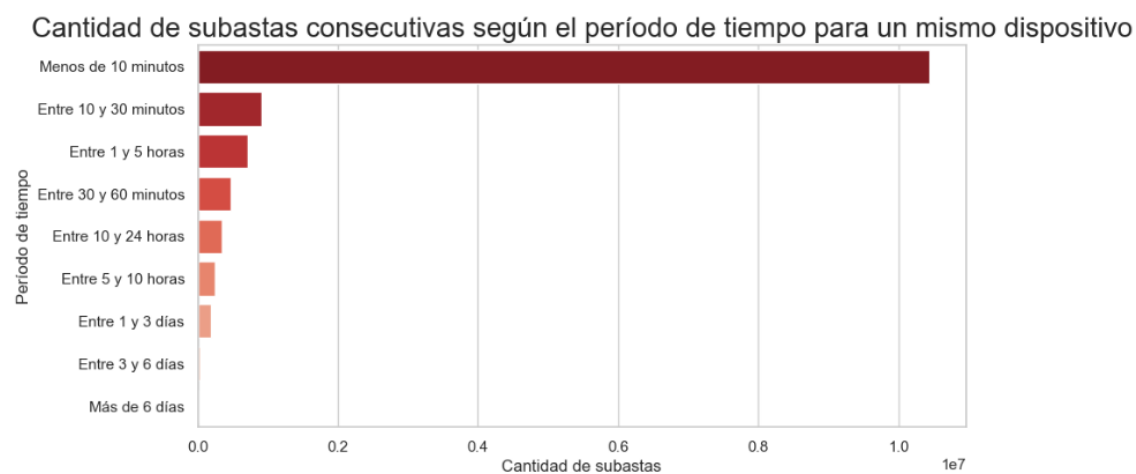


Figura 8: Subastas consecutivas en las que participa un mismo dispositivo.

A partir del gráfico de la Figura 8 podemos decir:

- La mayor cantidad de subastas consecutivas en las que participa un mismo dispositivo se da en menos de 10 minutos, con una cantidad bastante considerable. Esto nos puede pasar por dos opciones principalmente:
 - Que esta clara diferencia sea por aquellos usuarios en los que figura que participan casi continuamente (con una diferencia de segundos nomás, lo que nos parece que los datos están erróneos).
 - Que los usuarios utilizan el dispositivo ésta cantidad de tiempo, es decir, el tiempo seguido en que un usuario juega o utiliza una aplicación. También se puede ver que los siguientes valores son *entre 10 y 30 minutos* y luego *entre 1 y 5 horas*.
- Podemos ver que a medida que pasan los días, la cantidad de subastas en las que participan, disminuyen fuertemente.

4.1.7. Existe alguna tendencia sobre la fuente de donde procede la subasta

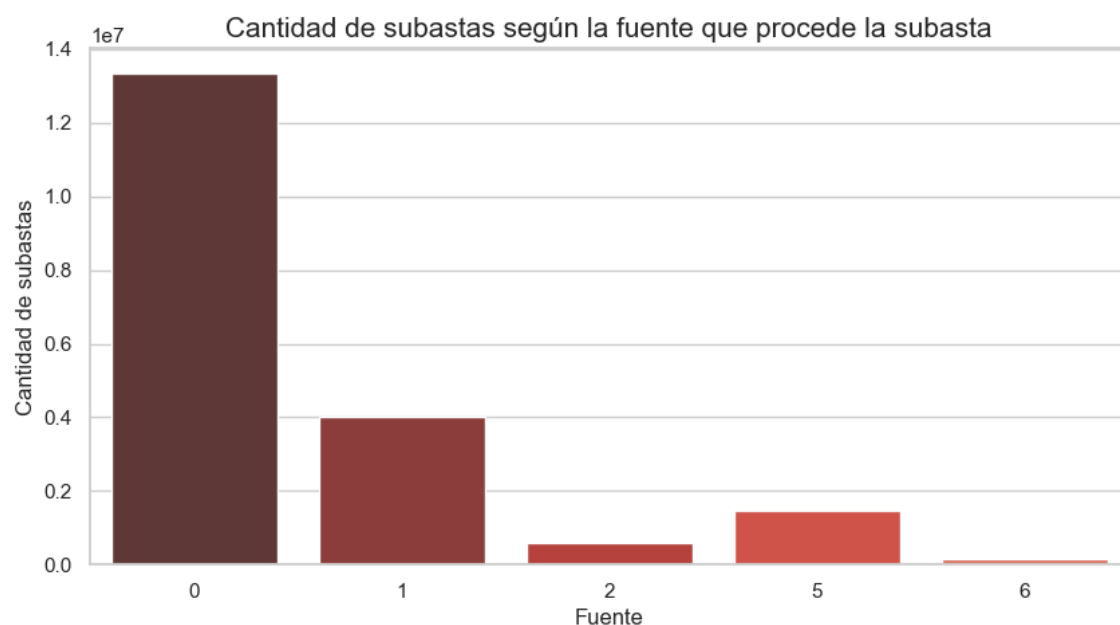


Figura 9: Cantidad de subastas según la fuente.

Si analizamos las subastas respecto a la fuente de la cual procede, podemos ver que hay una diferencia muy notoria entre que la más frecuente y las demás. Dicha fuente se representa con un 0, lamentablemente no tenemos información adicional para poder arriesgar o suponer de que valor realmente se trata.

4.2. Clicks

En esta sección se hará un análisis de los datos presentados en el archivo *“clicks.csv”*, el cual contiene información referida a todos los clicks ocurridos sobre las diferentes publicidades.

4.2.1. Cómo evoluciona en el tiempo la cantidad de clicks

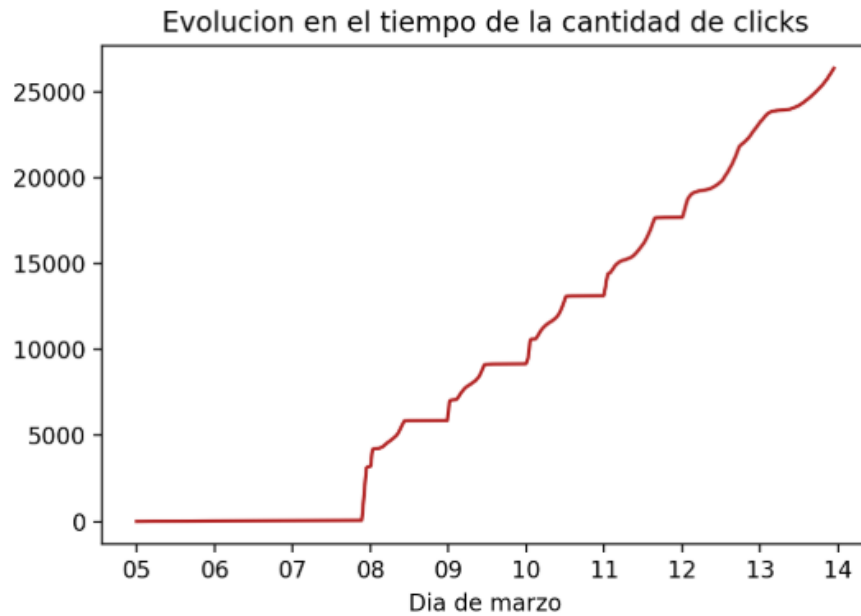


Figura 10: Cantidad de clicks registrados según el día.

Los primeros 3 días (del 5 al 7 de Marzo), vemos que casi no hay clicks. Así que estudiamos esto en detalle:

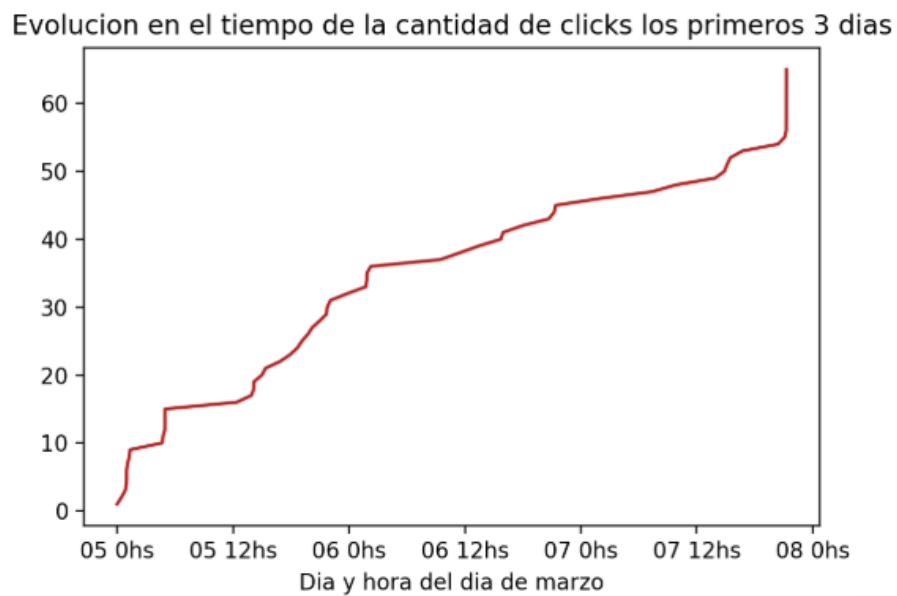


Figura 11: Cantidad de clicks registrados para los primero 3 días.

Se ve que los primeros 3 días hay apenas más de 50 clicks ¿Será esto un error del set de datos?

4.2.2. A qué hora se realizan la mayor cantidad de clicks

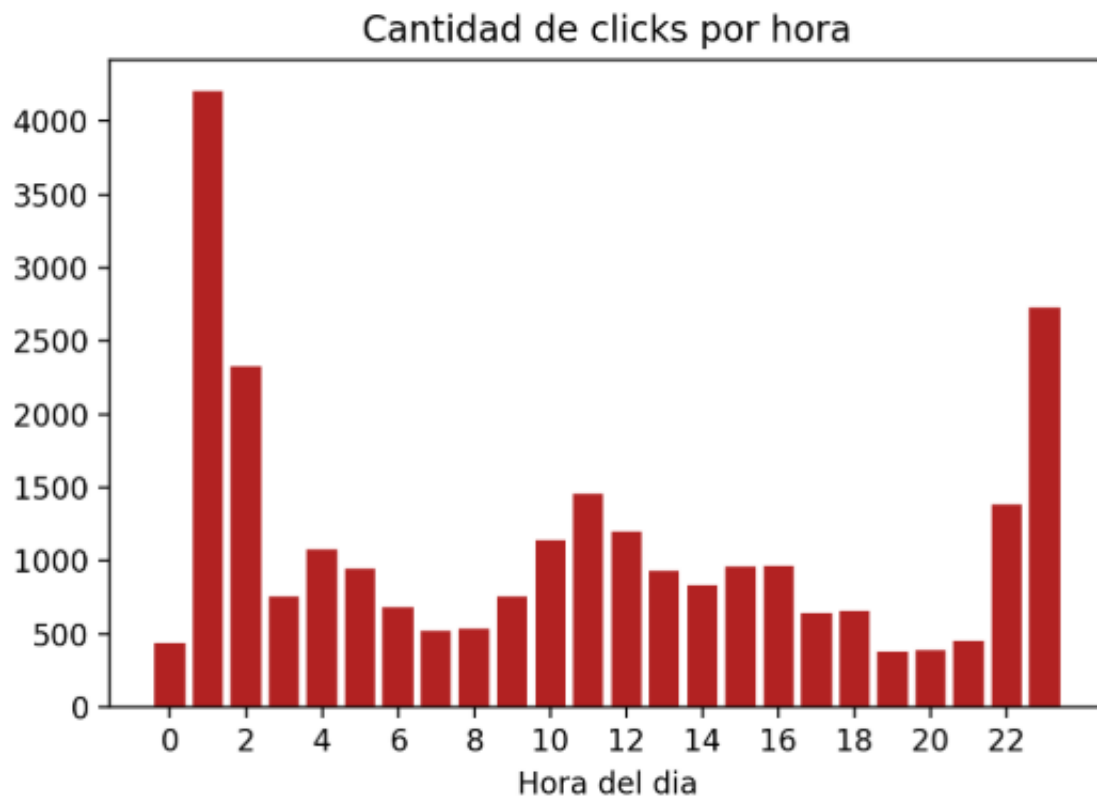


Figura 12: Cantidad de clicks registrados por hora.

Se ve que la mayor cantidad de clicks se realizan a la 1, con un pequeño pico a las 11hs. Por otro lado, es llamativo que a las 0hs hay muy pocos clicks dado que a las 23 y 1hs se encuentran la mayor cantidad de clicks.

4.2.3. En que parte de la pantalla los usuarios hacen click

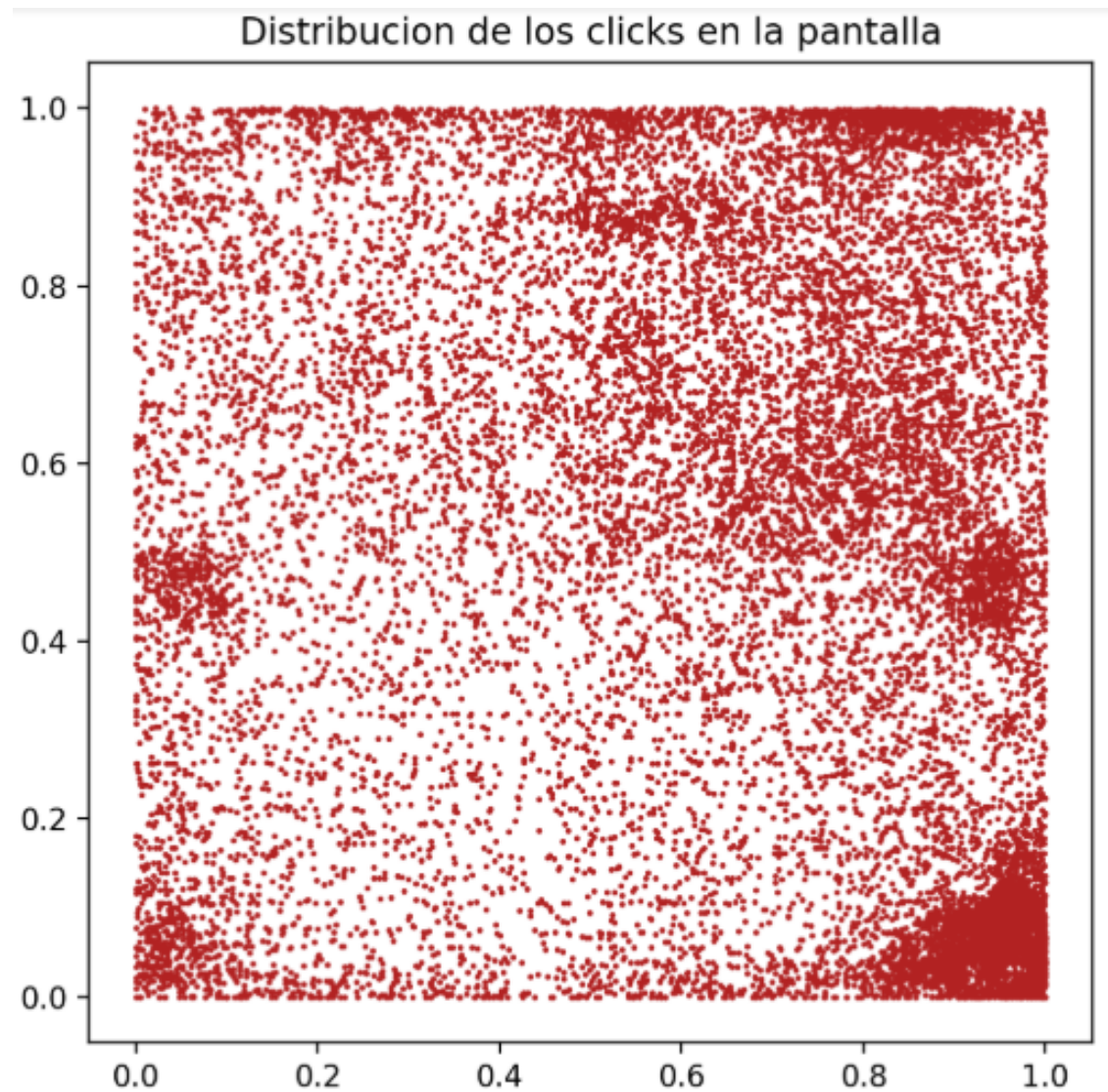


Figura 13: Distribución en la pantalla.

Se ve que la mayor concentración de los clicks están en la esquina inferior izquierda de la pantalla. Los bordes en general reciben más clicks que el centro de la pantalla. Cabe aclarar que se eliminaron muchos valores por tener una ubicación en Y mayor a 1.

4.2.4. Cuáles compañías telefónicas son las más usadas



Figura 14: Cantidad de clicks del Top 10 de compañías telefónicas.

Podemos ver una marcada diferencia entre la compañía con más clicks y las demás empresas telefónicas. Los otros puestos del podio (segundo y tercer lugar) se ve una pareja cantidad de clicks registrados.

4.2.5. Cuanto tardan los usuarios en hacer click



Figura 15: Histograma según el tiempo hasta hacer click.

Hay unos pocos valores mayores a 2500. El grueso de los valores está entre 0 y 200, así que analizamos en ese intervalo:

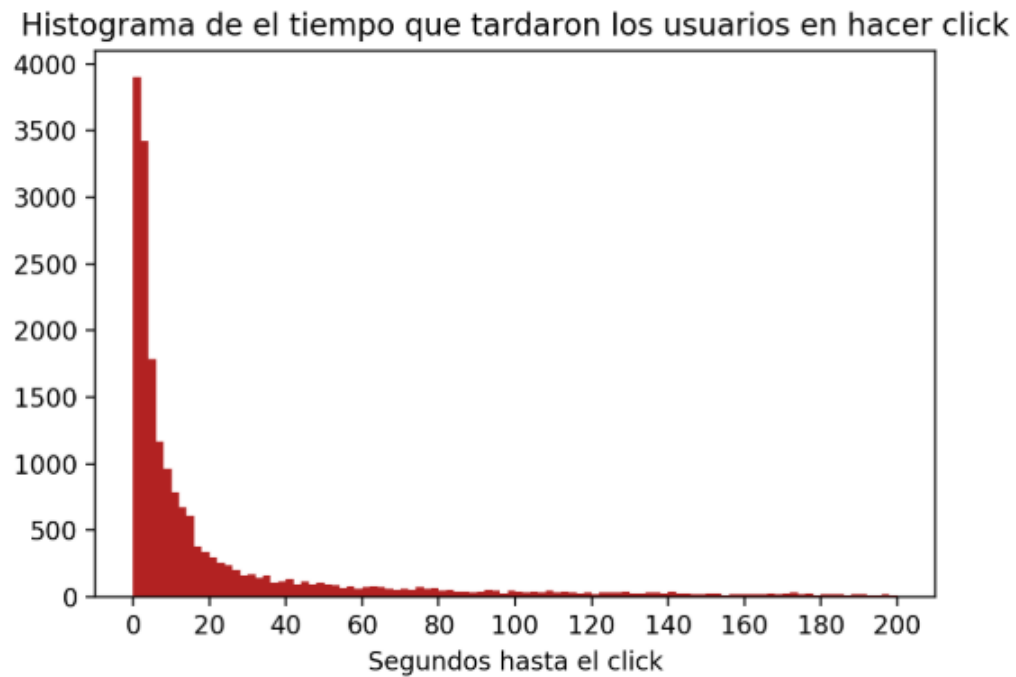


Figura 16: Histograma según el tiempo hasta hacer click.

De vuelta, podemos seguir acortando el intervalo para analizar la distribución de los tiempos más frecuentes.

Histograma de el tiempo que tardaron los usuarios en hacer click

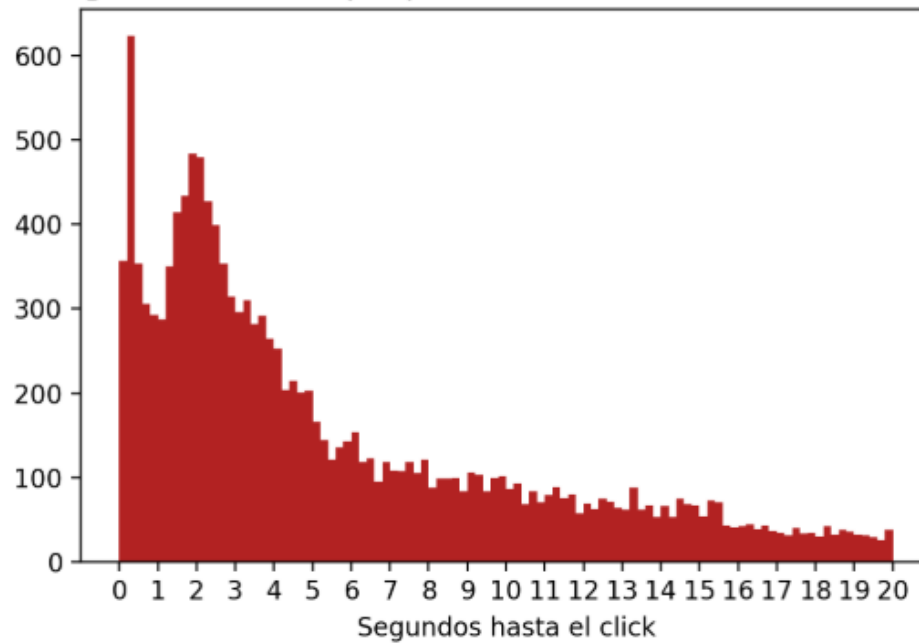


Figura 17: Histograma según el tiempo hasta hacer click.

Histograma de el tiempo que tardaron los usuarios en hacer click

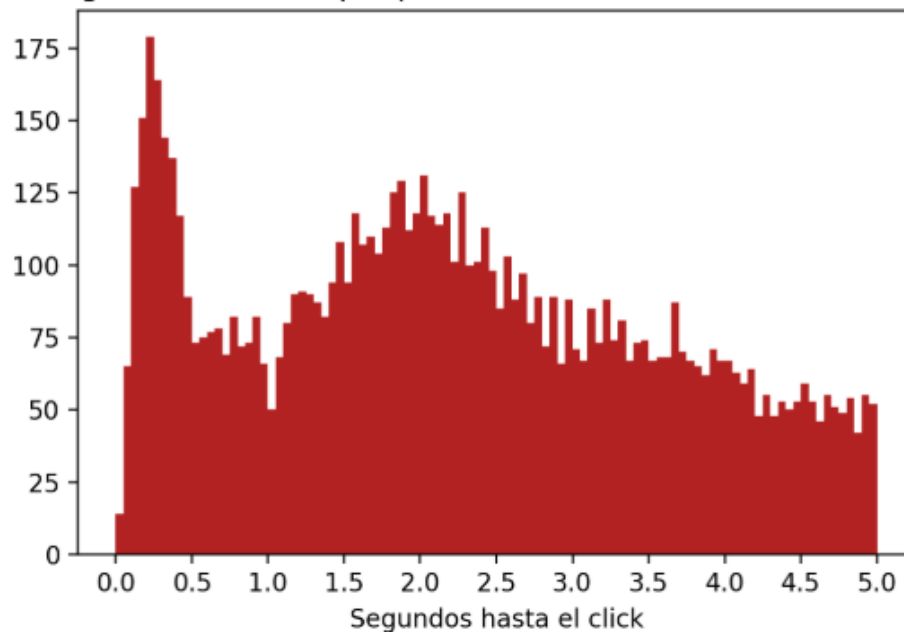


Figura 18: Histograma según el tiempo hasta hacer click.

Se ve que los valores más frecuentes están entre 0 y 0.5. A los usuarios no les lleva mucho

tiempo decidir.

4.2.6. Que marcas de celular registran más clicks

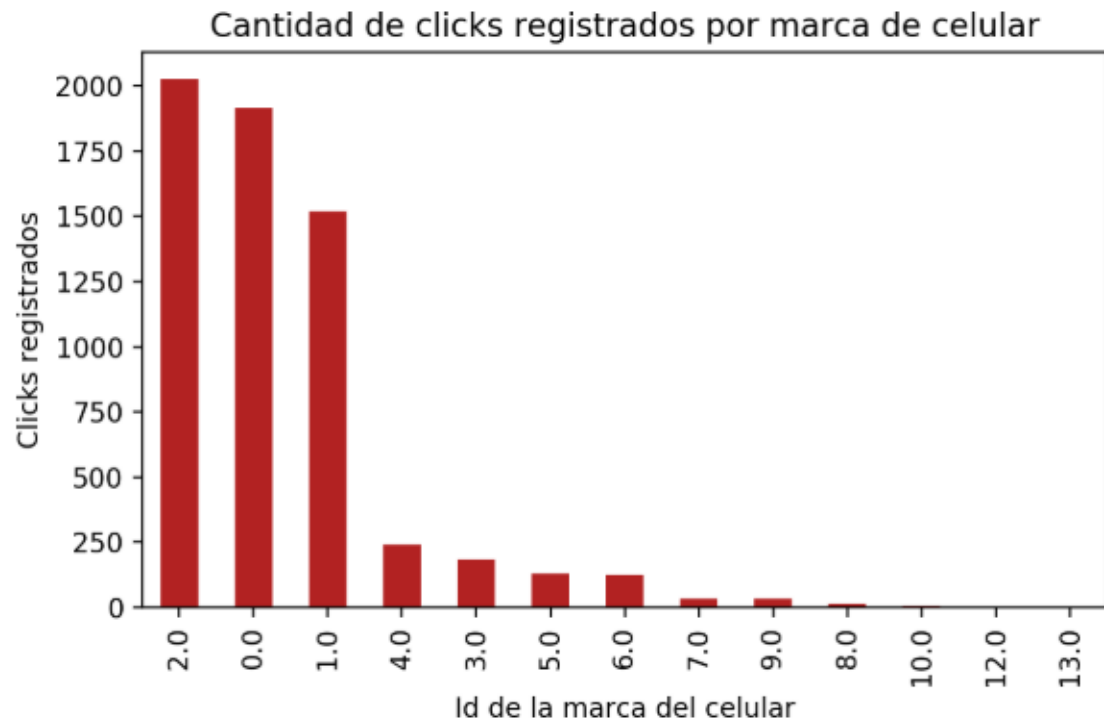


Figura 19: Cantidad de clicks registrados por marca de celular.

4.2.7. Que usuarios realizan mayor cantidad de clicks

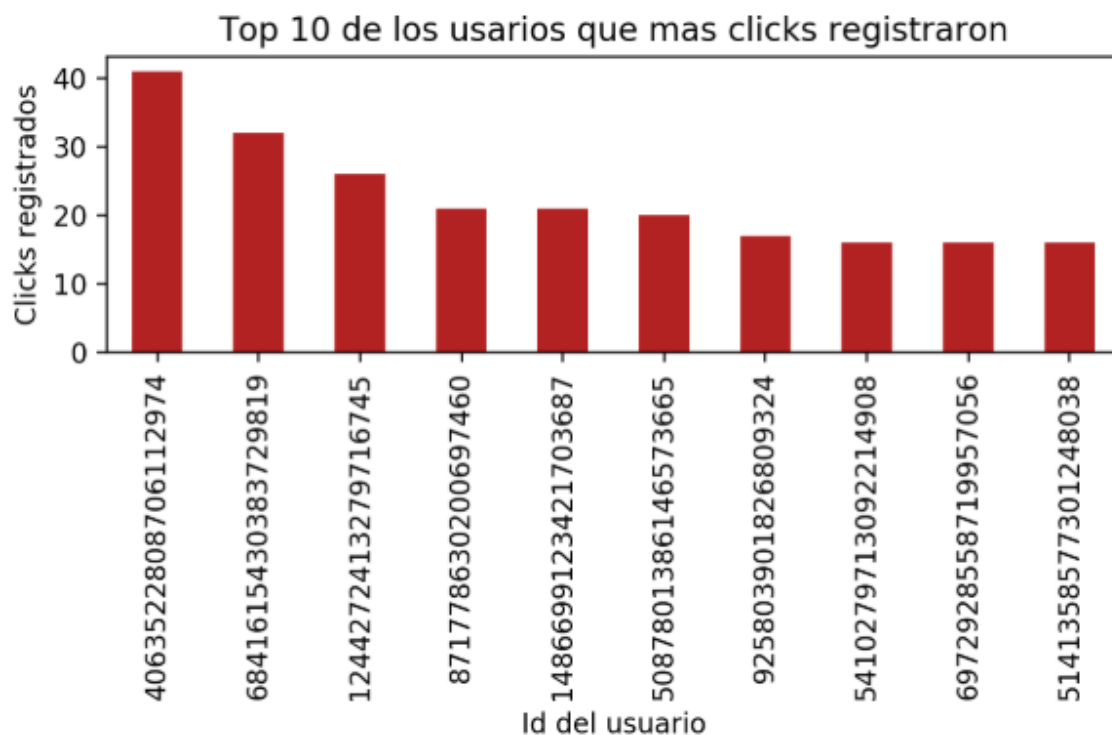


Figura 20: Cantidad de clicks registrados por ID del usuario del dispositivo.

Además, la media de cantidad de clicks por usuario es 1.56

4.2.8. Otros datos

Hay otros datos interesantes para analizar. Por ejemplo, si los clicks se hicieron mientras el usuario estaba conectado al wifi. Lamentablemente, todos los clicks dan falso para ese dato. Suponemos que hubo algún error al computar eso.

La latitud y la longitud están distorcionados, así que no se puede sacar ninguna conclusión interesante de esto.

Para el `advertiser_id`, existe un advertiser que tiene el 99.67% de los clicks. Lamentablemente no podemos saber cuál es ese advertiser.

4.3. Events

En dicha sección se realizará un análisis respecto a las características de aquellos eventos que se encuentran en `events.csv`. Con la denominación de `.evento`, hacemos referencia a cualquier tipo de acción categorizada dentro de una aplicación.

4.3.1. Cómo varía la cantidad de eventos acorde a las variables temporales



Figura 21: Cantidad de eventos registrados según la hora del día.

Como observamos en el gráfico de la Figura 21, la cantidad de eventos en función de la hora tiene una tendencia muy marcada: los mayores registros se dan entre las 22hs y las 2hs de la madrugada del día posterior. Luego de dicho horario, el tráfico disminuye fuertemente, lo que probablemente se deba a que en general, se corresponde con el horario de descanso. Y luego va aumentando a lo largo del día.



Figura 22: Cantidad de eventos registrados según el día.

Como mencionamos antes, el periodo estudiado es entre el 5 de Marzo y el 13 de Marzo de 2019. Como podemos ver, el tráfico fue creciendo gradualmente. No podemos saber a qué se debe pero cabe la posibilidad de que Jampp haya empleado alguna estrategia de marketing durante este periodo.

De la Figura 22 nos llama la atención la depresión del día 11 de Marzo, un factor posible es que corresponde al día Lunes.

4.3.2. Cómo varía según el tipo y características de evento

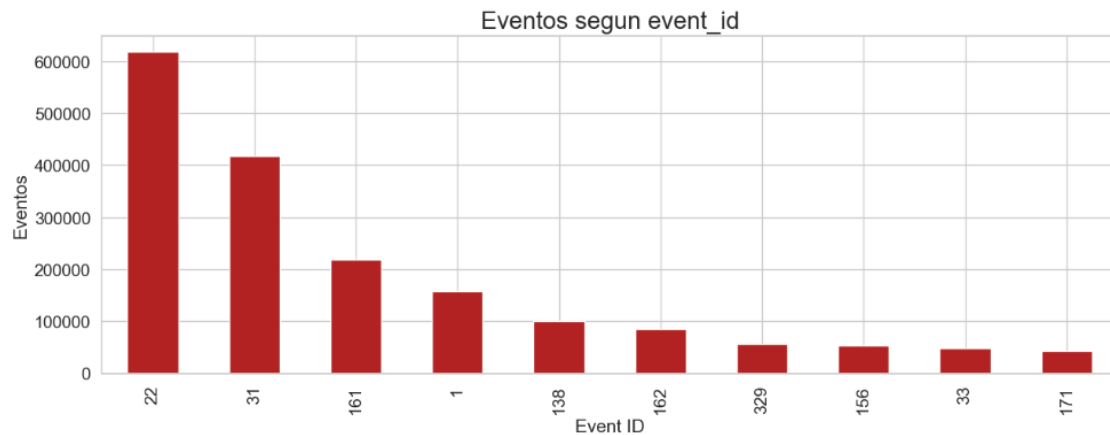


Figura 23: Cantidad de eventos registrados según el tipo de evento.

Estos son los 10 tipos de eventos que más veces ocurrieron, no queremos arriesgarnos a suponer de que se trata cada uno de ellos, pero podemos ver una notable diferencia entre los primeros dos tipos de eventos y los posteriores.

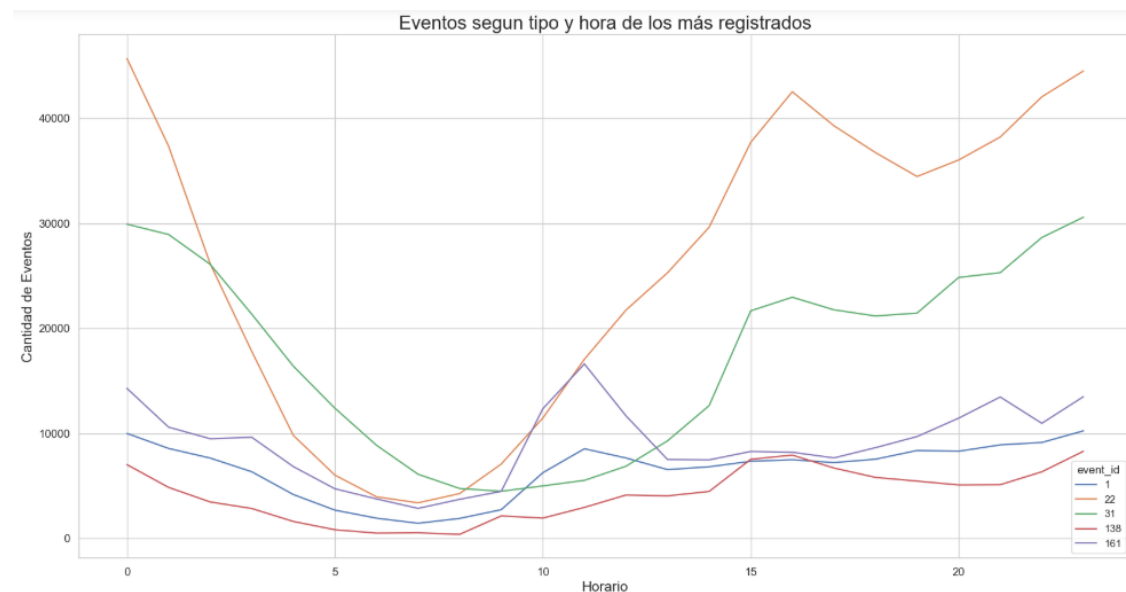


Figura 24: Cantidad de eventos registrados según el tipo de evento y la hora del día.

Como podemos observar, se refuerza la conclusión de la Figura 21 en las que hay un decaimiento en los eventos entre las 5hs y las 10hs, aunque la pendiente de cada tipo de evento es de distinta forma. También podemos observar que hay un marcado pico de eventos 1 y 161 a las 11 hs que no se corresponde con los demás eventos.

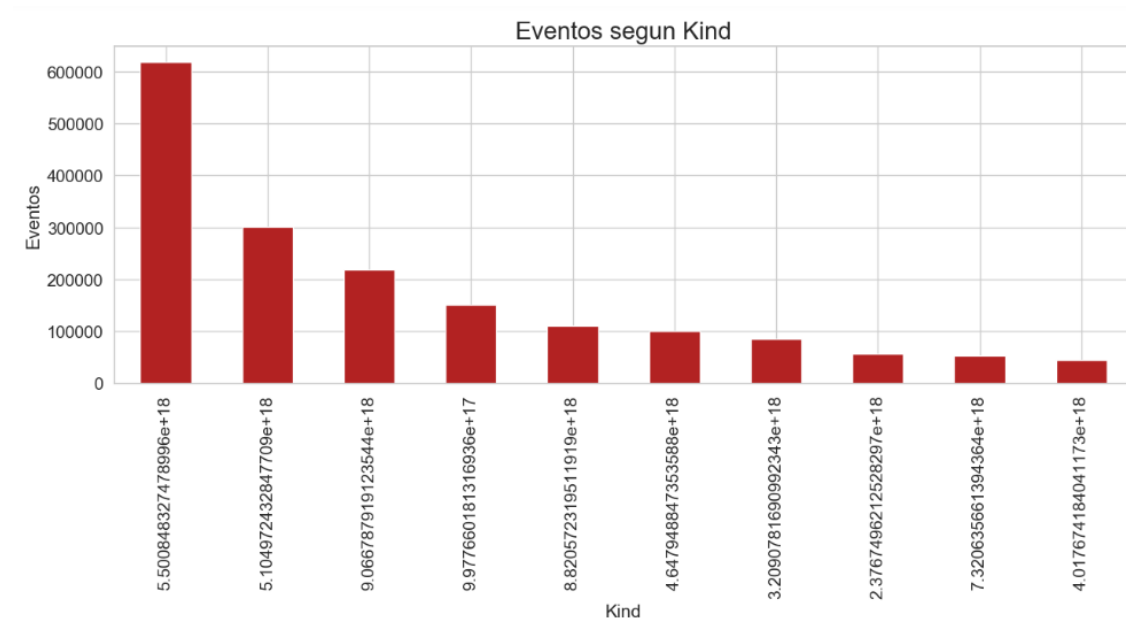


Figura 25: Cantidad de eventos registrados según Kind.

Como se ve en el gráfico, hay un kind dominante, con otros tres que tienen cierta relevancia, y luego varios más que se podrían considerar despreciables.

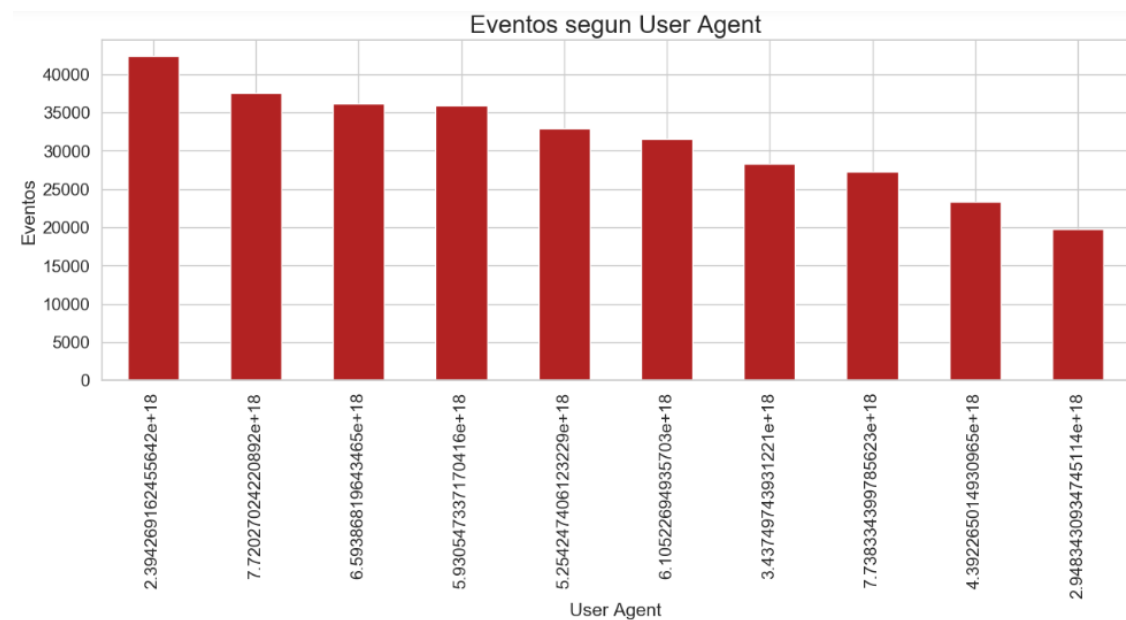


Figura 26: Cantidad de eventos registrados según User Agent.

Observando la Figura 26, podemos ver que los user agent están distribuidos de una forma más uniforme que el resto de las variables.

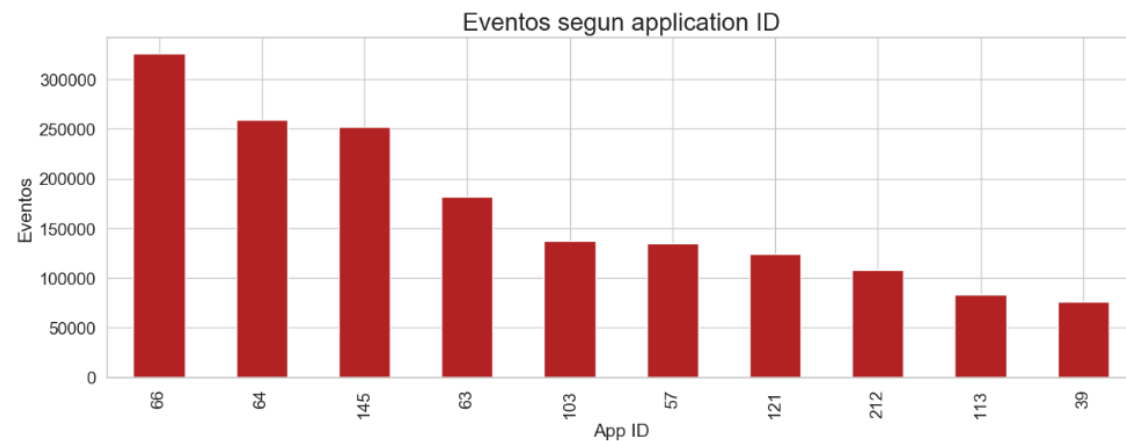


Figura 27: Cantidad de eventos registrados según Application Id.

Vemos que hay varias apps que acumulan mucho trafico, si tuvieramos que arriesgarnos diríamos que esas son las rede sociales como facebook, instagram, o twitter, entre otras apps de uso masivo.

4.3.3. Cómo varía la cantidad de eventos acorde a las variables geográficas

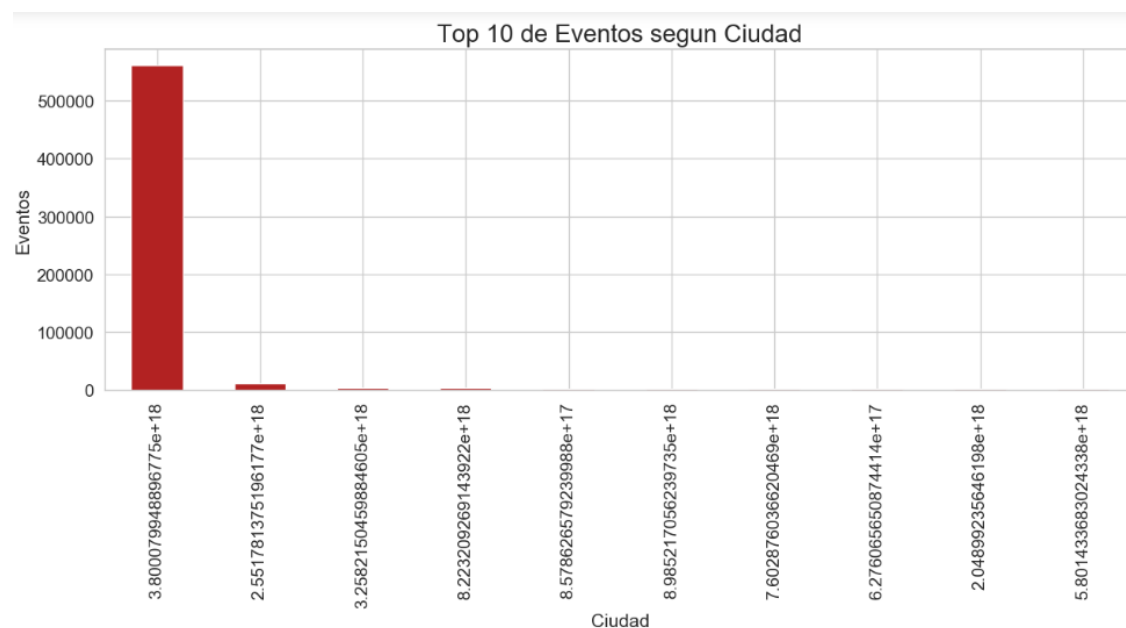


Figura 28: Cantidad de eventos registrados según la Ciudad.

Como se puede observar, casi el total del tráfico se origina en la misma ciudad y vuelve a las demás despreciables, si tuviéramos que arriesgarnos diríamos que se trata de la capital de Uruguay, Montevideo.

4.3.4. Cómo varía según las características del dispositivo

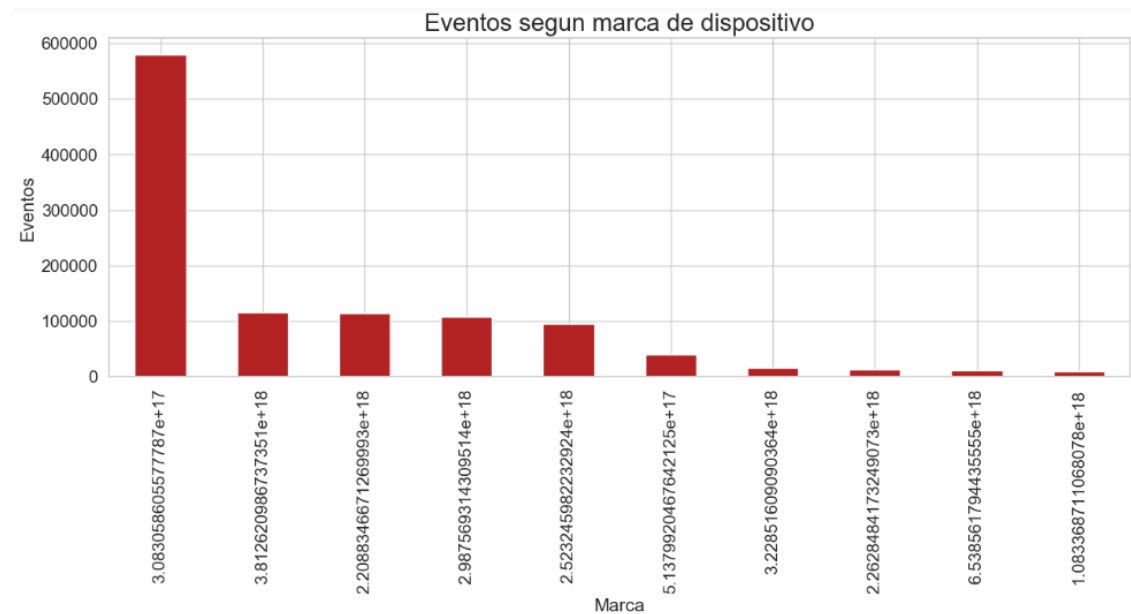


Figura 29: Cantidad de eventos registrados según la marca del dispositivo.

Observamos que hay una marca de dispositivo que opaca a las demás, no queremos arriesgarnos a decir de cuál se trata.

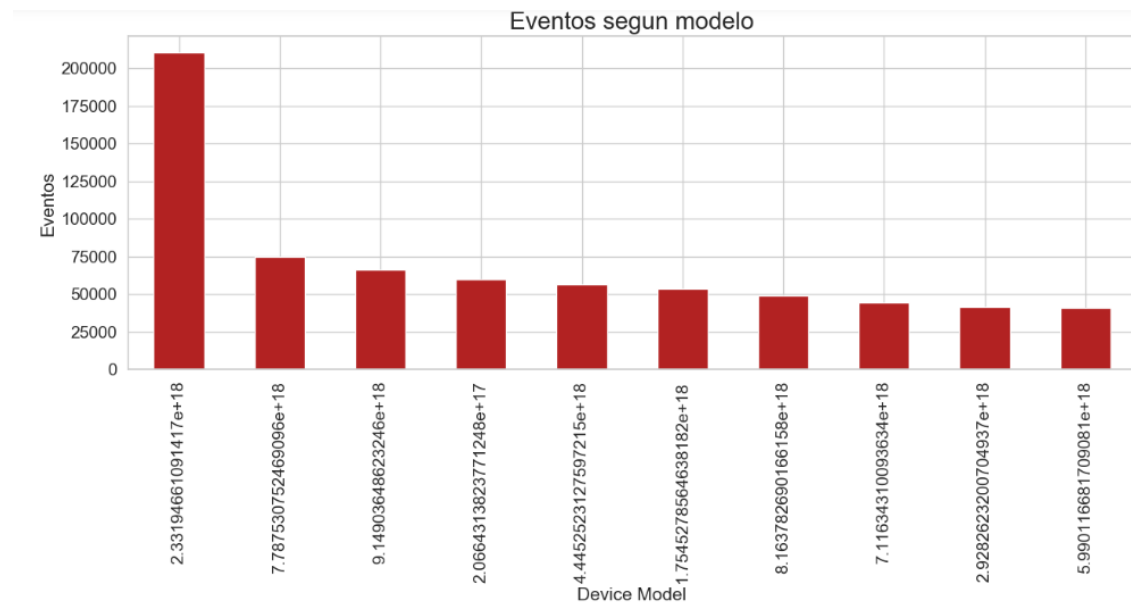


Figura 30: Cantidad de eventos registrados según el modelo del dispositivo.

Al igual que en con los datos referidos a la marca del dispositivo (Figura 29), vemos que hay un modelo de dispositivo que supera fuertemente a las demás, pero tampoco podríamos decir de

cuál se trata.

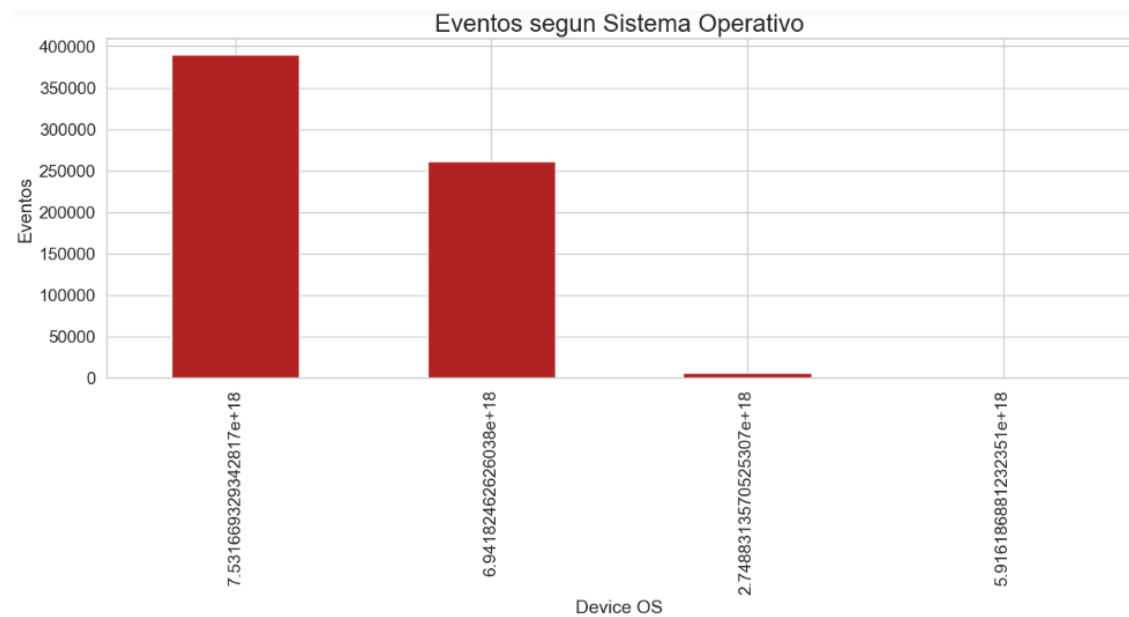


Figura 31: Cantidad de eventos registrados según el sistema operativo.

Observamos que hay dos sistemas operativos diferentes que acumulan la gran mayoría del tráfico, y otros dos que acumulan muy poco.

4.3.5. Cómo varía según tipo de conexión

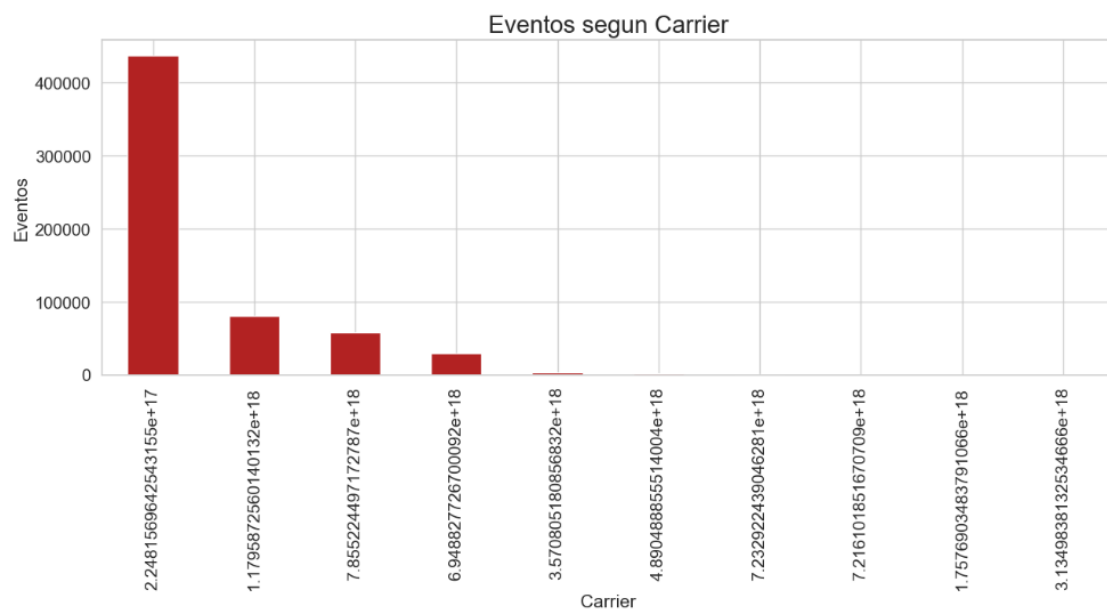


Figura 32: Cantidad de eventos registrados según Carrier.

Como vemos, hay un carrier (proveedor de telefonía móvil) que abarca la mayoría del tráfico y luego varios carriers con mucho menor tráfico.

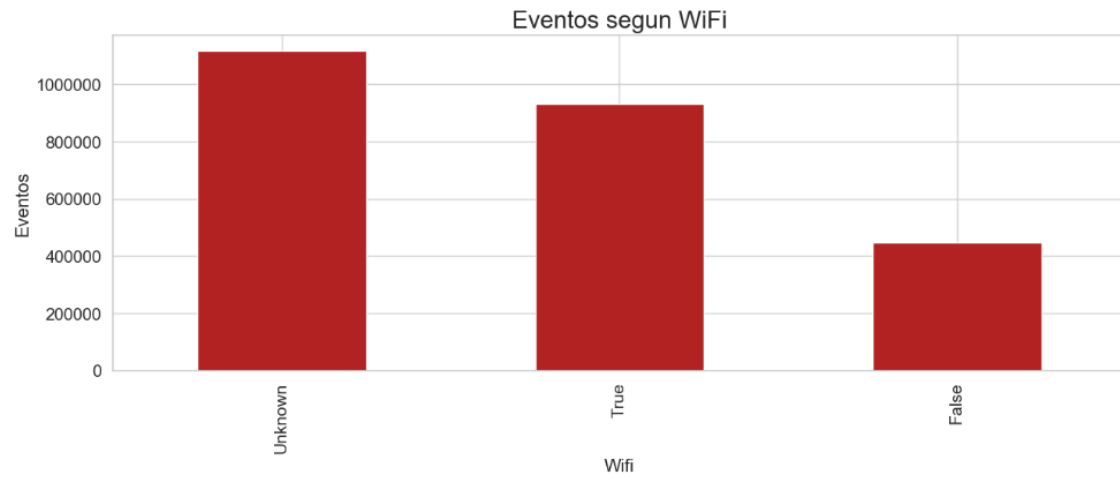


Figura 33: Cantidad de eventos registrados según conexión WiFi.

Observamos que el tráfico es mayor para la gente que accede vía Wifi.

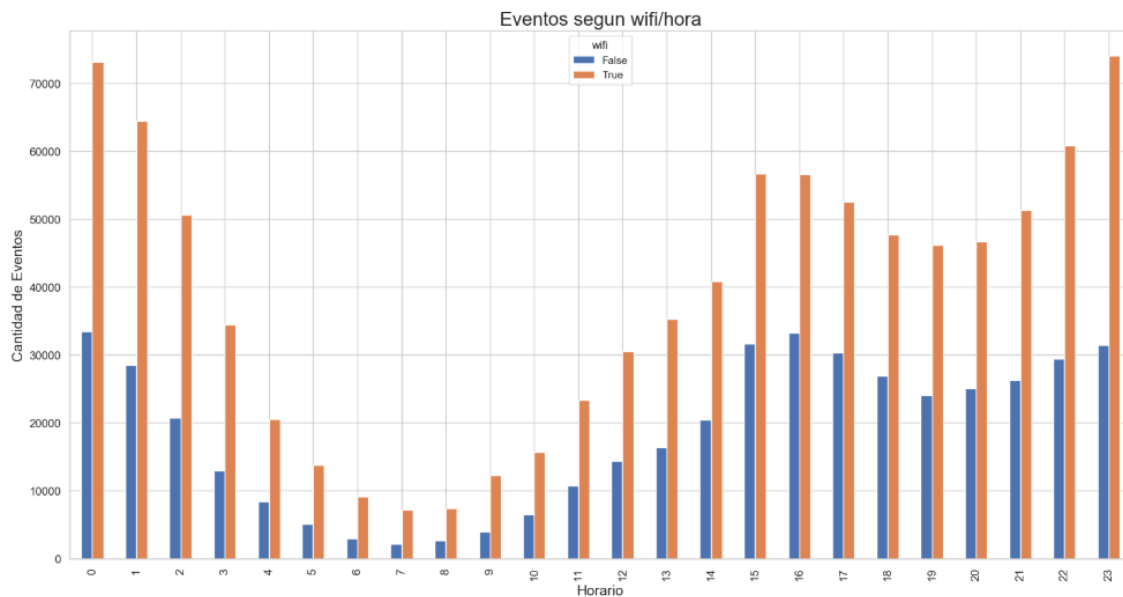


Figura 34: Cantidad de eventos registrados según conexión WiFi y hora del día.

Como podemos ver, se mantiene una tendencia similar para el tráfico independientemente de si la conexión es vía Wifi, o no lo es.

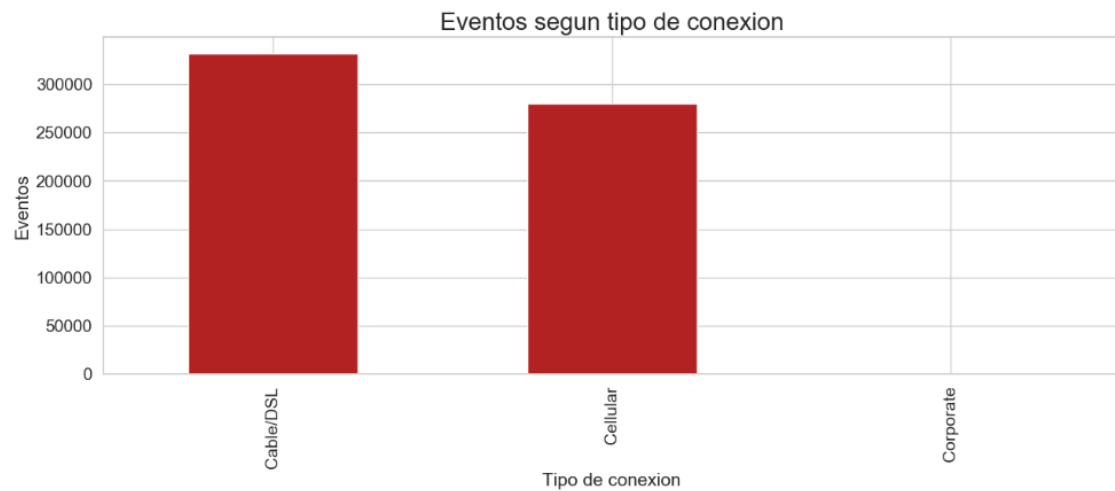


Figura 35: Cantidad de eventos registrados según el tipo de conexión.

Observamos que los tipos de conexión son Cable/DSL o Cellular, existe otra más, Corporate, pero es despreciable en relación a las otras dos.

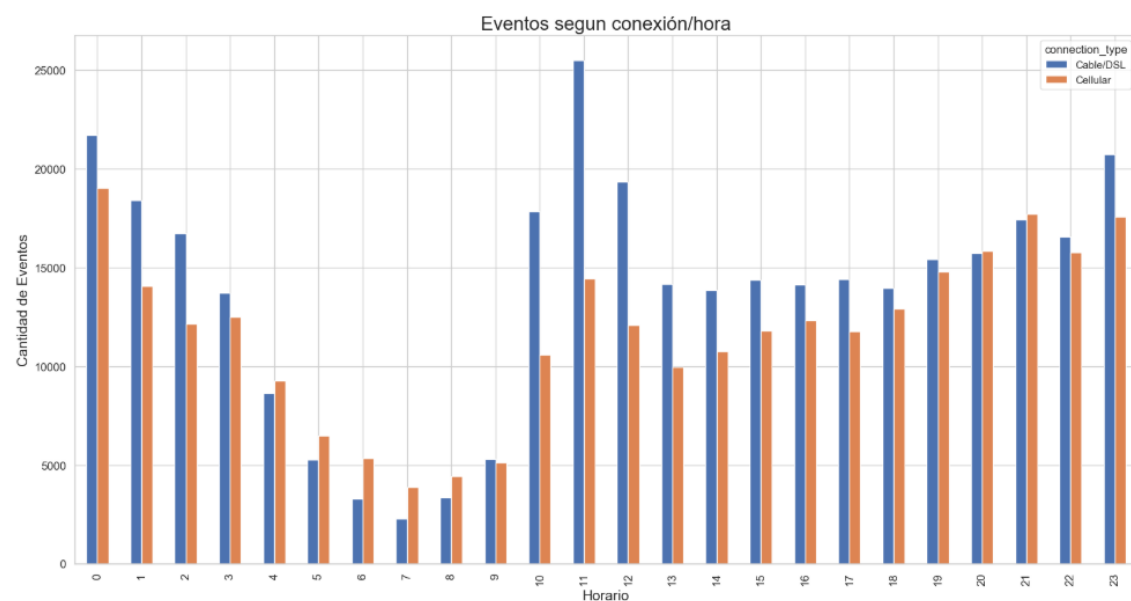


Figura 36: Cantidad de eventos registrados según el tipo de conexión y hora del día.

Se puede observar que el tráfico con tipo de conexión Cellular es menor que el de Cable, salvo a las 4,5,6,7,8,20 y 21 horas. La explicación que podríamos intentar dar es que en la franja temprana, de 4 a 7, la gente se encuentra viajando hacia sus lugares de trabajo, y los eventos ocurren en dispositivos móviles, con red Cellular (red móvil). No podemos suponer a qué se debe el aumento del tráfico vía Cellular a las 20 y 21 hs.

4.3.6. Cómo varía el lapso de tiempo entre dos eventos consecutivos para el mismo usuario

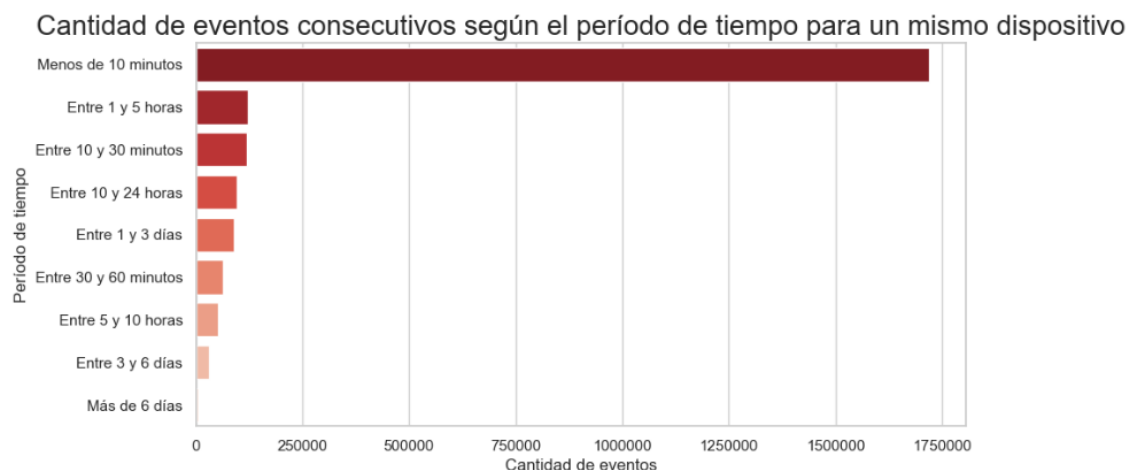


Figura 37: Cantidad de eventos consecutivos según el período de tiempo para un mismo dispositivo.

A partir de éste gráfico podemos decir:

- La mayor cantidad de eventos generado por un mismo dispositivo se da cada menos de 10 minutos, con una cantidad bastante considerable. Esto nos puede pasar por dos opciones principalmente:
 - Que esta clara diferencia sea por aquellos usuarios en los generan continuamente.
 - Que los usuarios utilizan el dispositivo ésta cantidad de tiempo seguida, también se puede ver que los siguientes valores son *entre 10 y 30 minutos* y luego *entre 1 y 5 horas*.
- Podemos ver que a medida que pasan los días, la cantidad de eventos, disminuyen fuertemente.
- También nos llamó la atención la poca cantidad de eventos (comparados con los otros grupos), que se generan *entre 30 y 60 minutos*

4.4. Installs

Aquí se analizarán los datos presentes en el archivo *“installs.csv”* que hacen referencia a toda la información relacionada con la conversión de un dispositivo, es decir, cuando el usuario instala la aplicación de la cual se mostró una publicidad.

4.4.1. Cómo varía la cantidad de instalaciones acorde a los días

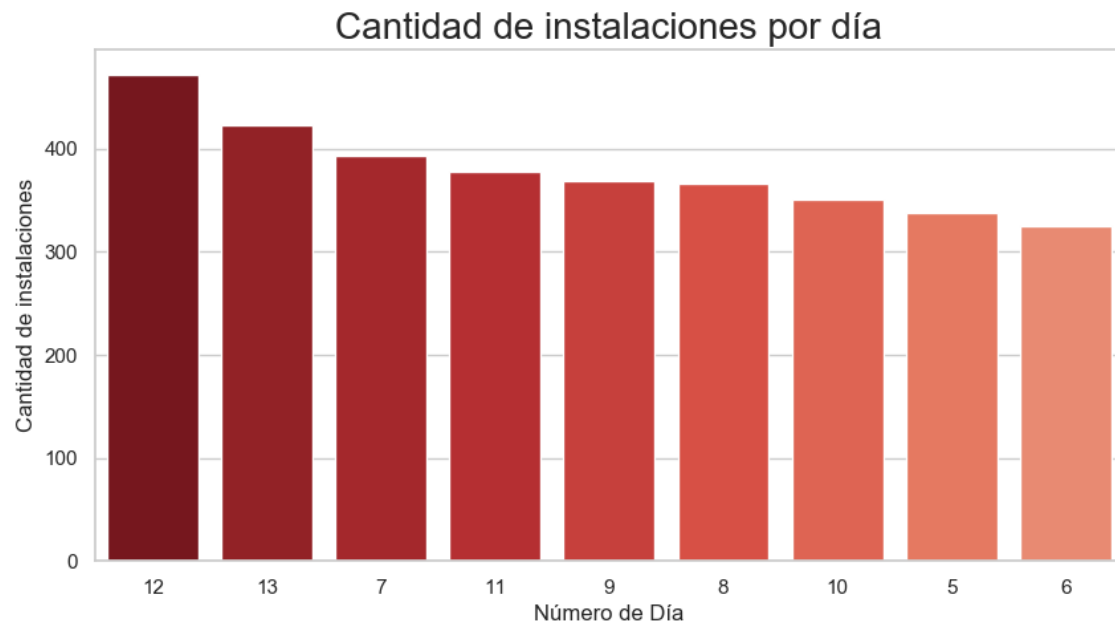


Figura 38: Cantidad de instalaciones según el día.

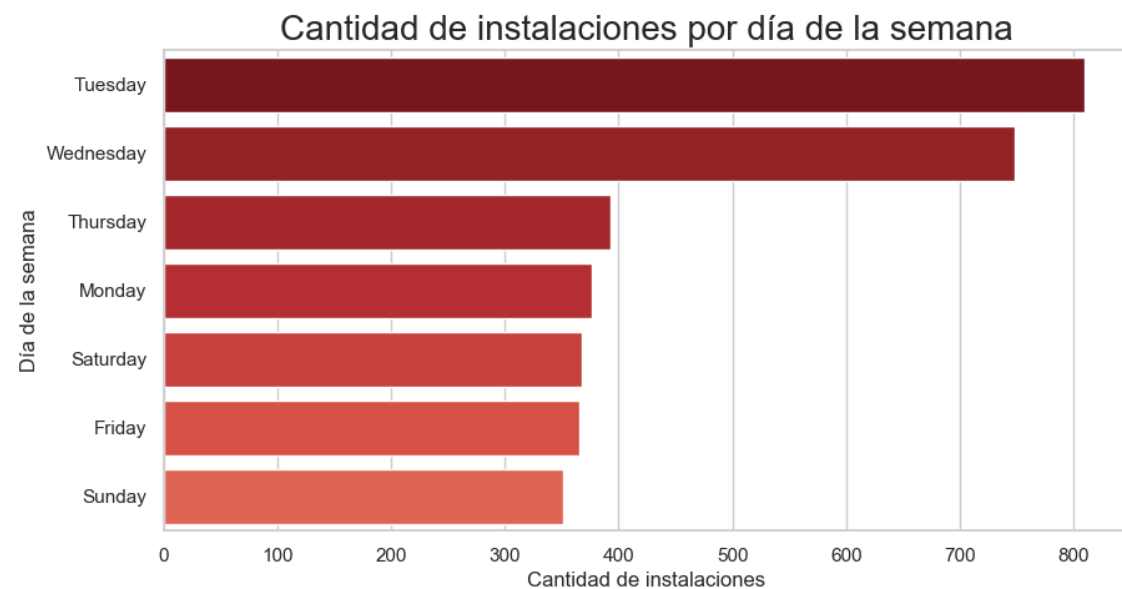


Figura 39: Cantidad de instalaciones según el día de la semana.

Observando y analizando los éstos dos últimos gráficos, podemos ver que la mayor cantidad de instalaciones se dan los días 12 (Martes) y 13 (Miércoles) de Marzo.

4.4.2. A qué hora se realizan la mayor cantidad de instalaciones

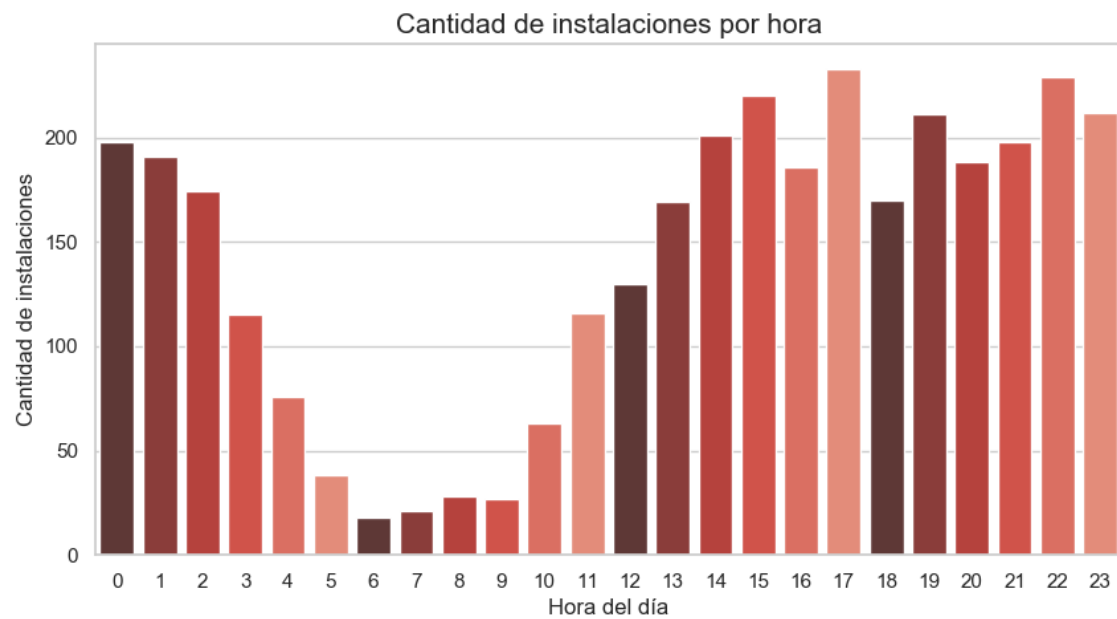


Figura 40: Cantidad de instalaciones según la hora del día.

Podemos observar que la mayor cantidad de instalaciones se realizaron a las 17hs, y luego a las 15hs. Nos llamó la atención como disminuye dicha cantidad en medio del intervalo, es decir, a las 16hs. Otro de los altos picos se produjo a las 22hs.

Por otra parte, la cantidad de instalaciones se reduce muy notablemente entre las 5am y las 9am.

4.4.3. Cómo se distribuye cantidad de instalaciones según el día y la hora

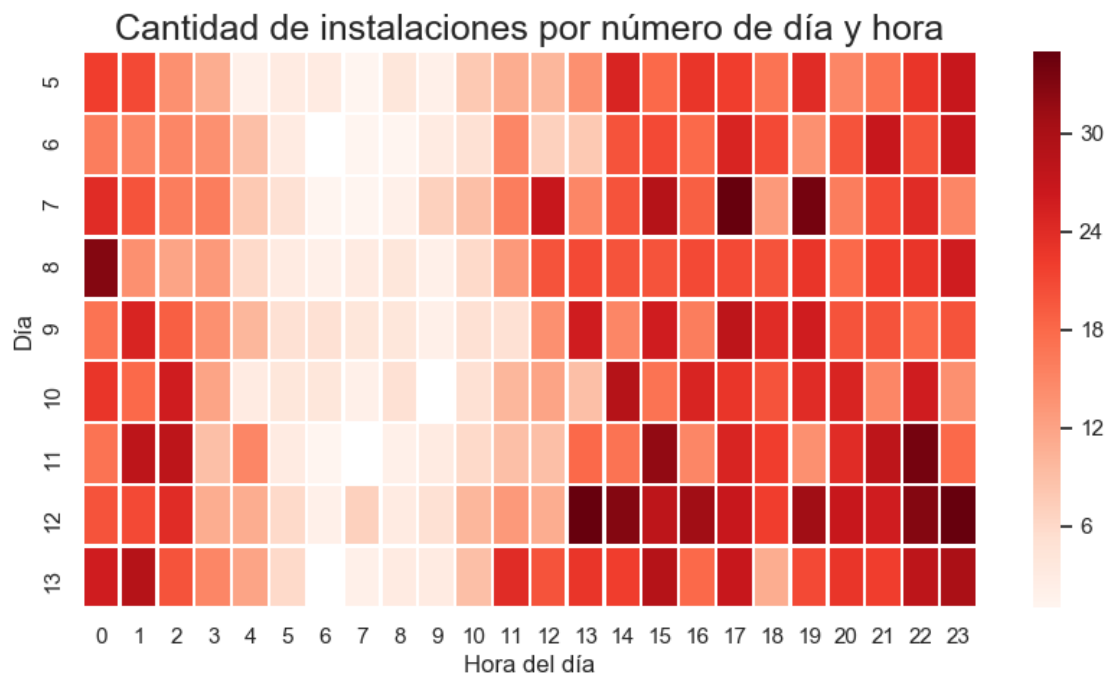


Figura 41: Instalaciones según la hora y el día.

Si analizamos de manera conjunta la cantidad de instalaciones por día y por el horario producido, podemos observar que se mantienen lo analizado para en las Figuras 38 y 40.

Por otra parte, vemos que para el día 7 de Marzo se realizaron muchas instalaciones a las 17hs y 19hs, habiendo una disminución considerable para las 18hs de dicho día.

4.4.4. Cómo varía según las características del dispositivo

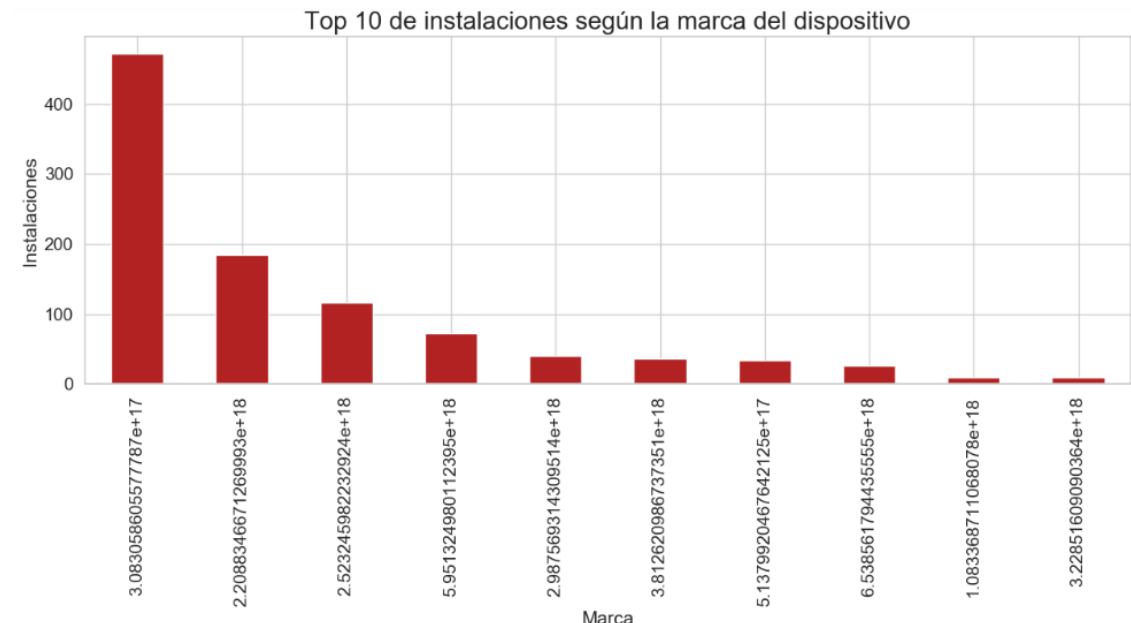


Figura 42: Cantidad de instalaciones según la marca.

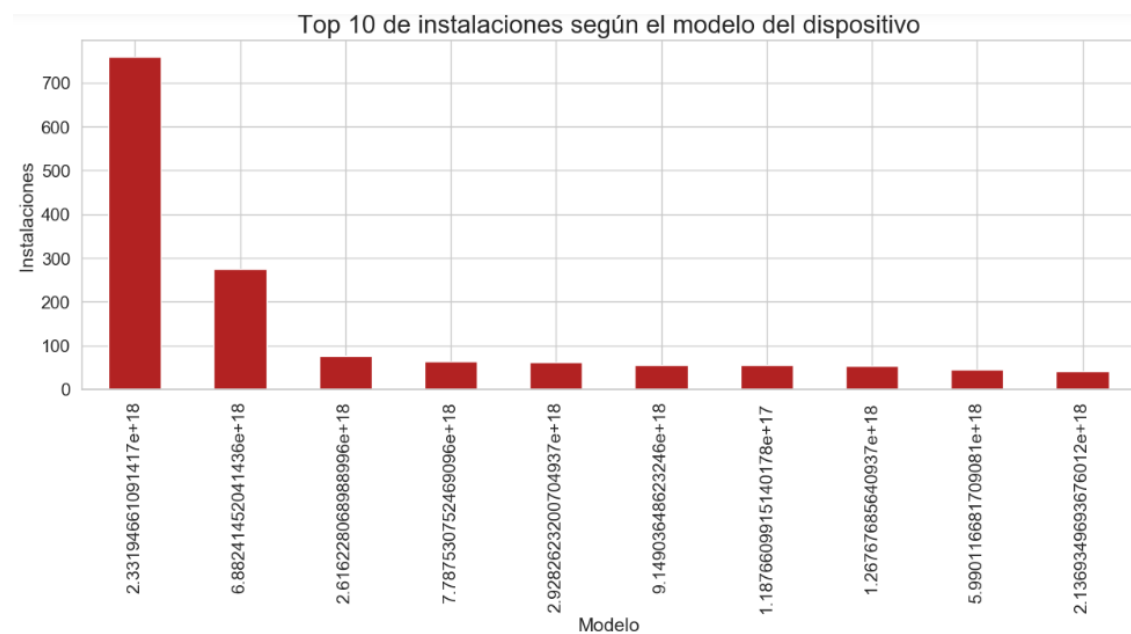


Figura 43: Instalaciones según el modelo del dispositivo utilizado.

Observando ambos gráficos, podemos ver que hay un valor que sobresalta con respecto a los demás, en los cuales la cantidad de instalaciones se vuelve muy menor. Lamentablemente no tenemos información adicional para poder determinar que marca y modelo representan los valores

correspondientes para poder sacar una mejor conclusión.

4.4.5. Cómo se distribuye la cantidad de instalaciones por usuario

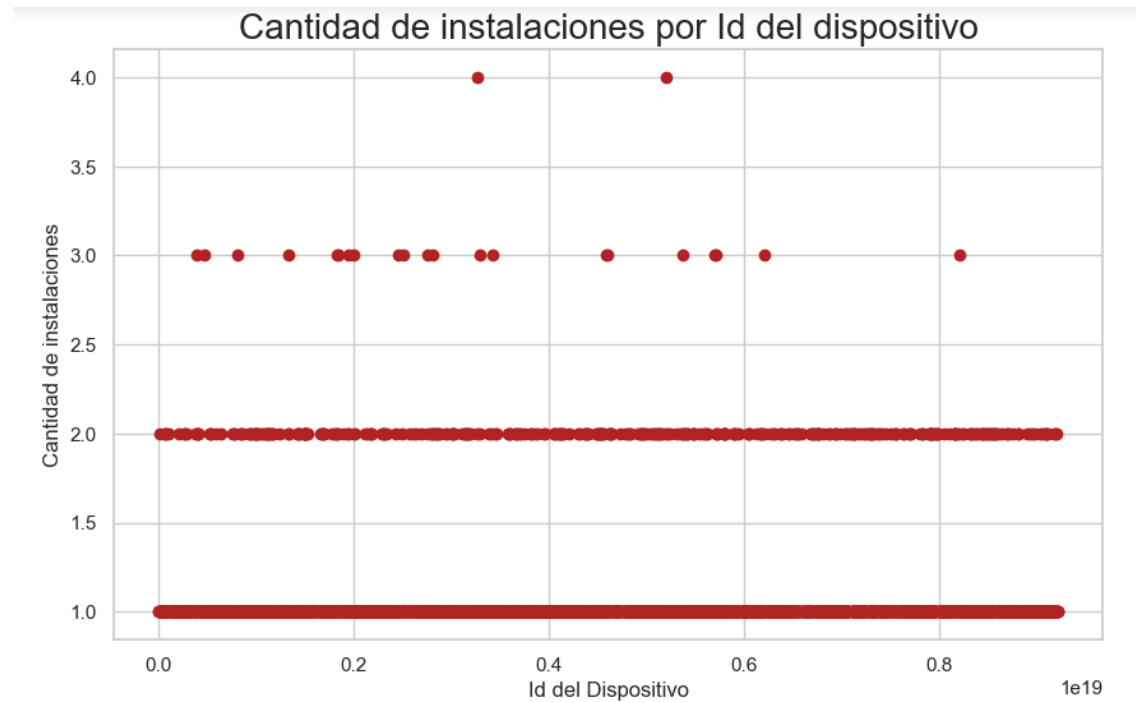


Figura 44: Instalaciones por dispositivo.

Viendo la cantidad de instalaciones por dispositivo podemos ver que tenemos muy pocos valores para 3 instalaciones, y solo dos valores para 4 instalaciones para el mismo usuario. Mientras que la mayor cantidad es de una instalación por dispositivo.

4.4.6. Cómo varía según tipo de conexión

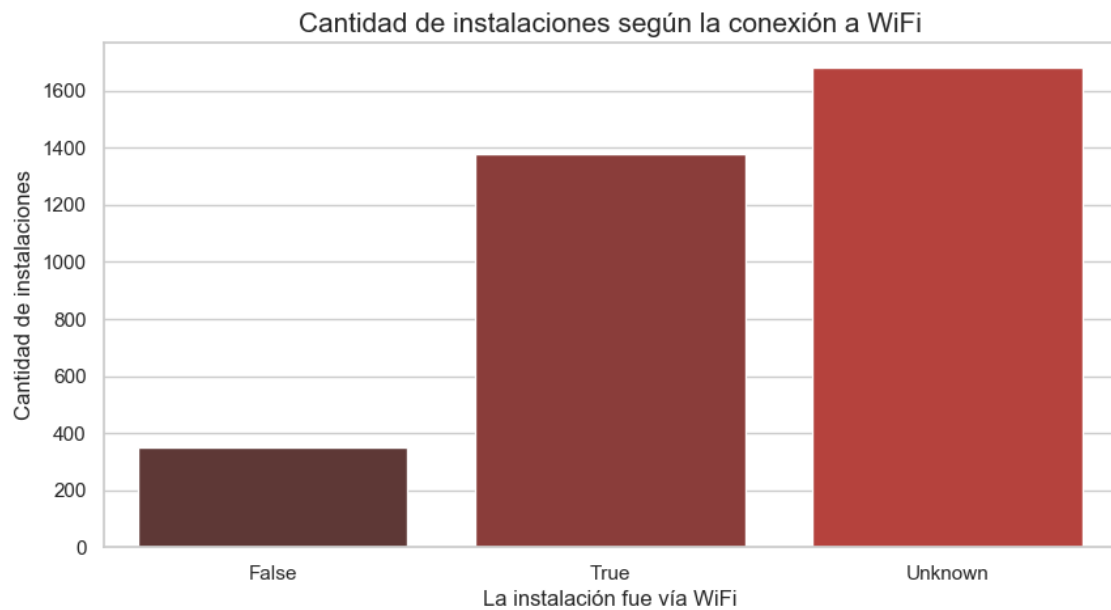


Figura 45: Instalaciones según conexión Wifi.

Aunque para este punto, tenemos varios valores nulos (*Desconocidos*), podemos ver que la mayor cantidad de instalaciones se da a través de una conexión WiFi.

4.4.7. Como se distribuyen las instalaciones según la aplicación

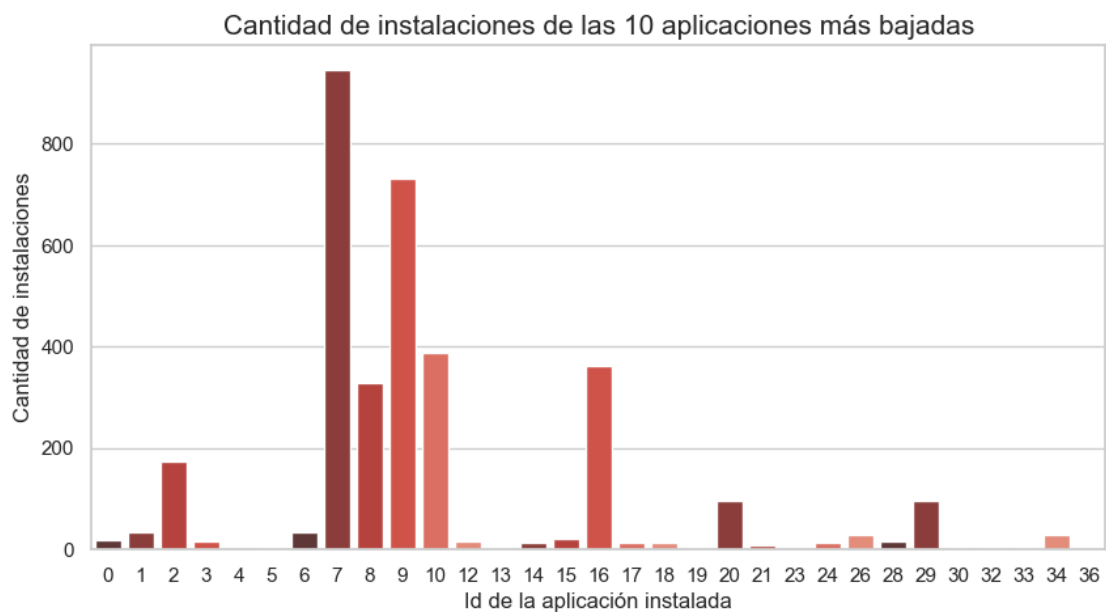


Figura 46: Cantidad de instalaciones según la aplicación.

Podemos ver que todo el flujo de instalaciones se dividen, principalmente, en dos aplicaciones. Ellas son la referenciada con un ID 7, y luego la 9. Si tendríamos que arriesgar diríamos que se trata de aplicaciones de delivery de comida online, como puede *PedidosYa* o *Glovo*, aunque ésta última no se dedica exclusivamente al delivery de comida.

4.4.8. Como varían las instalaciones según el anunciante

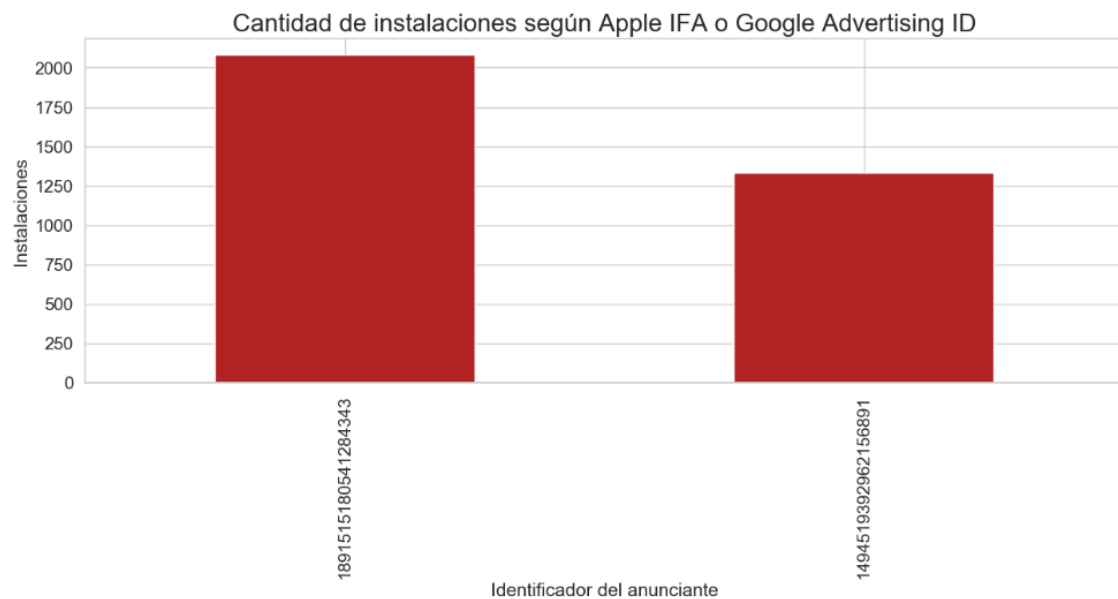


Figura 47: Cantidad de instalaciones según el anunciante.

Aquí vemos que hay un identificador del anunciante que tiene un poco menos de 1000 instalaciones más. El identificador puede ser *Apple IFA* o bien *Google Advertising ID*.

4.4.9. Existe alguna tendencia en la forma de instalación

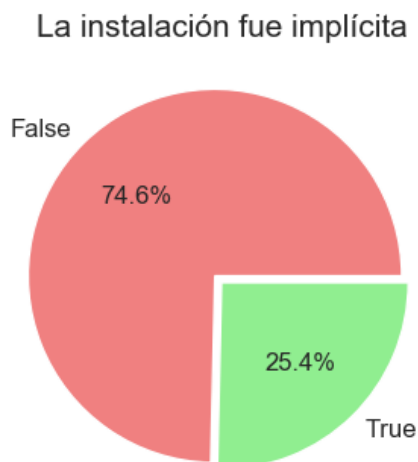


Figura 48: Verdadero si la instalación fue implícita. Falso en caso contrario.

Si vemos la forma en la que un usuario instaló la aplicación mostrada en una publicidad podemos ver que solo un poco más del 25 % decide hacerlo a través de la plataforma a seguir, es decir, que la mayoría no instala dicha aplicación siguiendo en enlace que genera la publicidad, sino que lo hace después por otros medios. Lamentablemente no tenemos información adicional sobre como se realizan aquellas instalaciones no implícitas.

5. Conclusión

Luego de la realización de este análisis del set de datos de la empresa *Jampp*, podemos resaltar y concluir:

- Efectivamente, la cantidad de instalaciones son mucho menos cantidad que los clicks o los eventos generados por los distintos usuarios. Vimos que solo dos usuarios realizaron 4 instalaciones. Siendo la media de 1.13 instalaciones.
- Con respecto a la hora del día en la cual se producen los distintos acontecimientos podemos decir que:
 - Como habíamos previsto, en la hora en la que la mayoría de los usuarios descansan (entre las 5am y 9am) tenemos un menor flujo de datos, como habíamos previsto.
 - Nos llamó mucho la atención que el mayor tráfico de datos ocurren en la noche y madrugada del día siguiente (salvo para instalaciones que tenemos un pico a las 17hs). Esto creemos que sucede que antes de acostarse, el usuario, decide distenderse un poco de lo que fue el día de trabajo/estudio y pasa un tiempo con el celular, ya sea para juegos o mirando sus redes sociales. También puede deberse a que para la cena se utilizan mucho más aquellas plataformas de delivery de comida online.
- El tiempo que pasa entre dos subastas consecutivas en las que participa un mismo dispositivo es de, con mayor frecuencia, menos de 10 minutos. Mientras que el siguiente valor corresponde al intervalo entre 10 y 30 minutos luego de la primera subasta. Esto creemos que ocurre debido a que los usuarios utilizan las distintas aplicaciones por un lapso de tiempo prolongado, lo que permite que, en dicho intervalo, se produzca la posibilidad de mostrarle más de una publicidad.
- Comparando los diferentes archivos, vemos que para el referido a las instalaciones, el día 7 de Marzo de 2019 se tiene una cantidad considerable de instalaciones. Mientras que, tanto para clicks, subastas y eventos, este día se encuentra dentro de los valores más bajos.
- Nos llamó mucho la atención que en el set de datos trabajados, ninguna instalación de las diferentes aplicaciones fue atribuida a Jampp. Lo que no sabemos si se trata de errores en dicho archivo, errores al no atribuirselas a Jampp, o que efectivamente, que ninguna instalación fue por la publicidad mostrada por la empresa.
- Analizando el tiempo en que los usuarios hacen click en el anuncio, podemos ver que los valores más frecuentes están entre 0 y 0.5 segundos. Por lo que a los distintos usuarios no les lleva mucho tiempo decidir.
- Dentro del intervalo de tiempo analizado, entre el 5 y el 13 de Marzo de 2019, podemos ver que los usuarios generan eventos con una media de entre 12 y 13 eventos (12.72 más precisamente). Lo que daría un promedio de un poco más de un evento por día.
- Para los diferentes acontecimientos, podemos ver que hay aplicaciones que reúnen la mayor cantidad de datos. Como también, hay marcas y modelos de dispositivos dominantes con respecto a los demás.

- Si observamos los eventos, podemos concluir que casi todo el flujo de datos se origina en la misma ciudad y vuelve a las demás despreciables, si tuviéramos que arriesgarnos diríamos que se trata de la capital de Uruguay, Montevideo. Ya que creemos que en ella la mayoría de la gente tiene acceso a los diferentes dispositivos móviles, y al ser una gran ciudad, pueden generar eventos para movilizarse, a través de aplicaciones que proporcionan vehículos de transporte con conductor (como *Uber* o *Cabify*), o utilizadas para delivery de comida online.
- Aunque para ver si el acontecimiento se realizó con conexión WiFi tenemos algunos valores en donde no hay datos referido a esto, igual podemos ver una clara diferencia entre las que utilizaron red WiFi con las que no.

Lamentablemente, como muchos datos sufrieron funciones de hash o corresponden a IDs internos de la empresa, no podemos determinar con un mayor grado de detalle las diferentes conclusiones. Como puede ser la aplicación más instalada, cual generó una mayor cantidad de eventos, entre otros acontecimientos.